5

Variation and Probability

Chapter Outline

Introduction	149
Models of variation	151
Probability matching	156
Trique trans-velar labial harmony	171
Comaltepec Chinantec tone alternation	180
High-tone behaviour in Zulu and elsewhere	190
Summary and conclusion	192
Doing phonology: Chiquihuitlan Mazatec	193

Introduction

When I discussed set notation during our discussion of 'phone books' and 'foam books' in Chapter 3, I wrote, 'Let's suppose for the moment that the word "foam" has only one pronunciation, whereas "phone" has … three … which are dependent on the context in which the word is found. While this is a simplification, for now, let's just suppose it's true'. We're now ready to discuss the nature of this simplification. Although I have not called your attention to it, perhaps you have realized that my phonetic transcriptions – intended as they are to capture the physical properties of a single particular *instance* or *token* of a word spoken by a particular speaker at a particular time – are actually highly idealized in the sense that I have not linked these transcriptions to a specific utterance by a specific speaker. In fact, in the history of the world, it is almost certainly the case that no two spoken utterances have ever been *exactly* alike. But the notation we have been employing up to now does not reflect this token-to-token variation at all. Instead, I have been using a single transcription that, at its very best, might represent either an *average* or an *idealised* pronunciation.

A far more representative display would involve a great number of transcriptions which all vary ever so slightly among one another in a manner that reflects the true variable nature of speech. Of course, even this sort of representation would not do genuine justice to reality, because we can never document the totality of realizations of any word. But we can, at least, employ a notational system that better approximates the true nature of speech. So look at the revised display in Figure 5.1.

The *clouds* or *pools* of tokens in this figure do a modicum of justice to the genuine variability of speech production by suggesting that every token differs slightly from every other token. The idea is that each token falls in its own location in some (as yet undefined) multi-dimensional articulatory/acoustic/auditory space. For example, the formant values of the vowel differ slightly from token to token; the tongue position of the nasal's oral closure is slightly different from token to token as well. Now, you'll immediately notice that I've transcribed every token within each pool in an identical way. So be it. This is just an impressionistic display and is not intended to convey the actual properties of the variation. For now at any rate, this sort of display suffices.

So, in addition to the phonetic distinctions found among alternants such as fõʊ̯̃m+b, fõʊ̯̃n, and fõʊ̯̃ŋ+k this other sort of phonetic *variation* shows that categories which are clearly defined *phonologically* are not so clearly defined *phonetically*. Phonological categories are clear-cut and discrete in that meaning is not *gradiently* affected by sound substitutions: sound substitutions either change meaning, merge meaning, or maintain meaning, and that's it. By contrast, gradually lowering the tongue in going from 'bid'

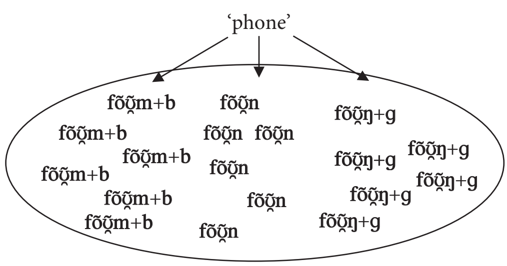

Figure 5.1 Variation within alternants.

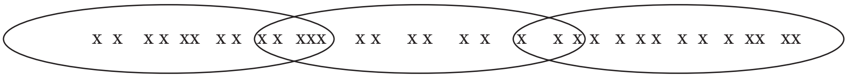

Figure 5.2 Variation on a single phonetic dimension.

to 'bed' to 'bad' does not, of course, produce corresponding intermediate *meanings* such as 'biddish bed' or 'beddish bad'. That is, the phonetic variation that we may observe among speech tokens has no direct correlates in terms of the categories to which these tokens belong: the gradience is only phonetic and never semantic. Indeed, as the linguist André Martinet remarked in 1975, 'Linguistic identity does not imply physical sameness ... Discreteness does not rule out infinite variety'. So, while all tokens within a category are identical in semantic terms, they are nonetheless phonetically diverse. Have a look at Figure 5.2.

Here, the 'x's represent tokens, the phonetic distinctions of which are suggested by their various locations along the one-dimensional scalar display. Moving from left to right, these tokens gradually change in terms of some phonetic property or other. Nonetheless, the semantic categories that learners come to impose on this gradient distribution are completely clear-cut and discrete. Despite the phonetic variation, some tokens fall into one semantic category, others fall into the other semantic categories.

But what about the intersecting regions of the sets? Certain tokens find themselves in two semantic categories, or at the boundary between one category and another. These tokens are ambiguous between one discrete category and another. (Later in this chapter, these tokens will be argued to play a very important role in the sound system.) But still, *phonetic* gradience has no correlate in terms of *category* gradience. It's certainly *not* the case that these tokens combine semantic elements of the two words, producing a *meaning* somewhere in between one category and another: 'discreteness in meaning does not rule out infinite variety in sound, and infinite variety in sound does not rule out discrete categoricity in meaning'.

Models of variation

Phonetic variation is inherent to speech production, but what might be the *origin* of this variation, and how did it come to so thoroughly permeate the system? We consider this question now, entertaining three plausible accounts.

One possibility is that phonetic constraints on speech are not so strictly imposed, and so speakers engage in an approximation of sorts. That is, their speech more or less resembles the speech around them, with a little deviation here, a little deviation there, that is somehow 'tolerated' at the cognitive level. We can refer to this approach as the *relaxed constraints model*.

Alternatively, there may indeed be strict cognitive constraints on speech. But if so, then what accounts for the phonetic variation that is undeniably present? There are two common proposals for the cognitive organization of this variation, often referred to as the *prototype model*, and the *exemplar model*, respectively. Both of these models (and also the relaxed constraint model) allow for the possibility that linguistic sound categories emerge through experience with individual examples or instances of perceptual events that come to self-organize themselves into distinct sets or categories. But these two approaches do have important differences between them. Proponents of the prototype model of categorization would propose that speakers have very exact internalized phonetic 'targets' for their speech (be these targets articulatory, acoustic, auditory, or some combination of these) but they don't hit these targets each and every time. This would be something like an expert dart player who inevitably misses the bull's eye at least once in a while: the darts are clustered around the target, but even an expert can't hope to get it just right every time.

Some version of the prototype model has been assumed by many linguists at least as far back as the nineteenth century. For example, in his book of 1880, the linguist Hermann Paul wrote, 'However much … movement may be the result of training … it still remains left to chance whether the pronunciation be uttered with absolute exactness, or whether slight deviation from the correct path towards one side or the other manifests itself'. While Paul does not specifically propose an abstract prototype, he nonetheless assumes that there is a single articulatory 'target' that speakers aim for. In the prototype model then, there is an exact, though abstract, category-defining value that emerges upon experience with individual tokens, and so any observed within-category variation is viewed as a deviation from this prototype.

What distinguishes the exemplar model is that perceptual categories are defined as the set of all experienced instances of the category, such that variation among tokens actually contributes to the categorial properties themselves. That is, a given category is the culmination of the variable forms themselves, in that the distribution of tokens is not viewed as a *deviation* but is instead viewed as a defining property of the category. So, within-category variation is thus part and parcel of the category itself. But what about the *origin* of variation under the exemplar approach? One idea is that speech patterns are copied again and again by generation after generation, variation included, but inevitably with very slight inexactitudes once in a while. According to the exemplar model, one generation's variation – minor chance inexactitudes included – serves as the next generation's template for copy. So variation may be viewed as the accumulation of very minor inexactitudes both within and between generations of speakers; the long-term product of accurate-though-imperfect copying of ambient speech patterns. The result is that tokens within a category cluster around each other, with each generation's distribution of tokens differing ever so slightly from both the preceding and following generations' tokens. As we'll see momentarily, these slight differences may come to play a significant role in the way sounds change over time. If we further assume, as is reasonable, that more recently encountered tokens leave a stronger memory trace than do more remote tokens, then we can further account for the sound changes observable even across the lifetime of a single speaker.

This approach to linguistic categorization is hardly new, having been proposed in the nineteenth century by Mikołaj Kruszewski. Discussing a hypothetical case of a slightly fronted **k** (which he writes *k'*), with variants k'_1, k'_2, k'_3, and so on, he wrote in 1883, 'Our characteristic, unconscious memory of the articulation of sound *k'* should be a complex recollection of all articulations of *k'* which we have performed. But not all of these articulations are arranged equally in the memory. For this reason, after performing the articulation of k'_3, the chances of performing k'_4 are much greater than they are for k'_1, etc.'

Hermann Paul, although, as noted, he does not seem to espouse an exemplar approach, nonetheless allows for his prototypes to evolve as a consequence of better remembering recent (versus remote) tokens:

> Variability of production, which remains unnoticed because of the narrow limits in which it moves, gives the key to our comprehension of the otherwise incomprehensible fact that a change of usage in the sounds of a language sets in and comes to its fulfillment without the least suspicion on the part of those in whom this change is being carried out … If the motory sensation were to remain always unchanged as a memory-picture, the insignificant deviations would always centre round the same point with the same maximum of distance … The later impressions always have a stronger influence than the earlier. It is thus impossible to co-ordinate the sensation with the average of all the impressions during the whole course of life; rather, the numerically-

> speaking inferior may, by the fact of their freshness, outbalance the weight of the more frequent … There thus gradually arises, by adding together all the displacements (which we can hardly imagine small enough) a notable difference …

The three approaches to speech variation are presented in the flowchart in Figure 5.3.

When we analyse the behaviour of an *individual*, it would seem extremely difficult to figure out which approach – relaxed constraint, exemplar, or prototype – is best at characterizing the origin of variation, and the nature of sound category formation. The semantic category itself is largely observable, since we can, at least in theory, investigate the phonetic properties of individual tokens, and also observe whether or not the tokens we are looking at correspond to a particular meaning. We can't directly observe the meaning that is associated with a given token of course, but we can probably determine if the speech signal was interpreted by a listener with the meaning intended by the speaker. In this way at least, meaning is observable. But since we are only dealing with the behaviour of the individual speaker without an interlocutor, there is really no reliable way to tease apart the three different approaches to variation and categorization that we are investigating. In short, all three approaches make untestable predictions about the categories and the variation characteristic of individual speakers. In Figure 5.4, the distribution of elements is exactly the same within each set, and so there is no empirical evidence favouring any one approach over the others.

It now becomes apparent that, under the prototype model, *all* variation must be regarded as mistaken and unintended. And since virtually every token deviates from the abstract prototype in some way, this means that virtually *all* speech is to a certain extent *mistaken* speech. In this sense at least, both the relaxed constraint model and the exemplar model have a

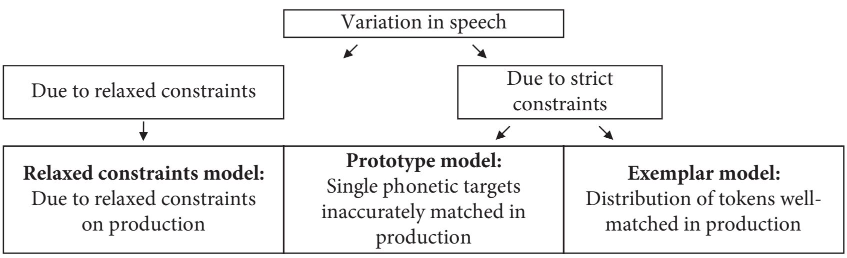

Figure 5.3 Three approaches to variation.

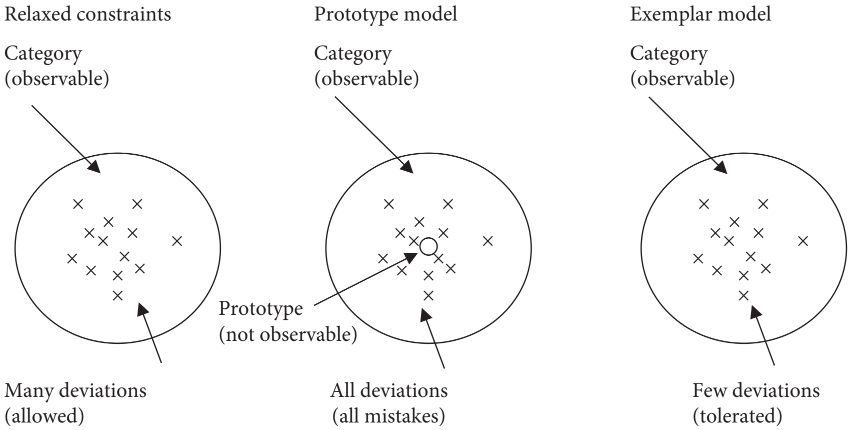

Figure 5.4 Again, three approaches to variation.

distinct philosophical advantage over the prototype model. After all, it's rather absurd to be forced into a theoretic corner in which virtually all speech is necessarily erroneous speech.

In accounting for the variation itself though, these latter two approaches are slightly different. The relaxed constraint model allows for some flexibility in the constraints on actual speech, and so all the variation is created anew by each speaker and each generation. Meanwhile, the exemplar model assumes a particularly tight match-up between the cognitive constraints and actual speech, since most of the variation is copied intact.

Now let's talk about the individual as both a *speaker* and a *listener*. By doing so, we are incorporating the social context in which this categorization procedure takes place: listeners are listening to other speakers, and speakers are speaking to other listeners. Therefore, we can now *compare* the speech of various speakers in order to determine the similarities and differences of their within-category variation. This comparison could, in theory, open a window into how listeners' categories are similar to or different from the categories of those they are listening to. Moreover, the *extent* of their similarity or difference might help determine which approach – the relaxed constraint model, the prototype model, or the exemplar model – is best at characterizing the categories and their variation. If variation is extremely well matched from speaker to speaker and from generation to generation, this would favour the exemplar model. This is because the exemplar model predicts very little difference in the nature and extent of variation from speaker to speaker and from generation to generation. By contrast, the other approaches allow for the possibility that the nature and extent of variation may be quite different from one speaker to another, and from one generation to the next, since according to these theories, variation is created anew by each speaker. We turn to this issue now, concluding that the exemplar model succeeds in accounting for the nature of this variation, whereas the relaxed constraints and prototype models do not fare as well.

Probability matching

Both speaker-to-speaker and generation-to-generation comparisons are very important to our understanding of sound categorization. These are the areas of (synchronic) *variation* and (diachronic) *sound change*, and it is here – as opposed to the investigation of individual speakers – where we might better understand the cognitive organization of the linguistic system. To be sure, investigating variation and sound change are indirect routes to understanding the nature of linguistic knowledge. But therein may lie their greatest advantages: by analysing variation and sound change, we make no presumptions about accessing the content of psychological states, which are inherently private and so unknowable to the outside world. Instead, we are comparing structural arrays of *genuine physical objects – speech tokens* – and so we harbour no illusions about the object of our inquiry. So the question we now turn to is this: what is the nature and extent of within-category *differences* in variation from speaker to speaker, and from generation to generation?

In order to answer this question, let's first talk a little bit about rats and ducks. It is well documented that animals appear to perform remarkably sophisticated statistical analyses as they navigate the world around them. For example, on the face of it, an animal foraging for food in the wild appears to be randomly searching high and low for a morsel here, a morsel there. However, it turns out that this behaviour is remarkably well matched in terms of actual payoff. What I mean is, the animal actually recapitulates the likelihood of payoff in terms of its foraging behaviour, spending more time in a patch of ground that has a greater payoff, and less time in a patch of ground that has lesser payoff. So, if two-thirds of the available food is in one region, and one-third is in another region, the animal very quickly comes to spend two-thirds of its foraging time in the one area, and only one-third of its foraging time in the other area. This phenomenon is known as *probability matching*.

A number of ingeniously simple experiments have been performed which show that animals indeed engage in probability matching. In one of the simplest studies, a rat placed in a T-Maze is rewarded with food 75 per cent of the time at one end, 25 per cent of the time at the other. When provided with feedback, the rat's foraging behaviour very quickly comes to match the probability of reward – running to the 'better' end 75 per cent of the time, the 'worse' end 25 per cent of the time. What's especially interesting is that the rat does not maximize its payoff. If it ran to the 75 per cent-payoff end 100 per cent of the time, it would be rewarded with food 75 per cent of the time. But by distributing its foraging in a way that matches the probability of payoff, it actually reduces its food intake. So, 75 per cent of the time it searches at the location where 75 per cent of the food is found, and 25 per cent of the time it searches at the location where 25 per cent of the food is. This means that that it only receives 61.5 per cent ($0.75 \times 0.75 + 0.25 \times 0.25$) of the total available food, as opposed to the maximum of 75 per cent.

As counter-intuitive as this result may seem, from a long-term, evolutionary perspective, the rat's behaviour makes very good sense indeed. Remember that this experiment only involves a single rat. But rats in the wild, of course, live in packs. If all rats were to forage only in the location with the greatest payoff, then fierce competition would result in a rapid depletion of resources. After these supplies run out, these rats might very well move on to the location with less food, and again compete fiercely for the rapidly diminishing resources. But consider a rat that bucks this strategy, and instead quickly matches its foraging behaviour to the probability of payoff. This rat would have less competition for resources at the location of lower payoff, guaranteeing itself a steadier intake of food. So those rats that engage in probability matching are in less competition for resources than those that forage exclusively in the locations of highest yield. Due to this reduced competition, these rats are more likely to survive and so transmit their foraging proclivities to their offspring, who, in turn, are more likely to survive. So probability matching benefits the individual, and, as a by-product, enhances the long-term stability and survival of the population as a whole. This behaviour, then, is the long-term emergent result of variable feeding strategies across individual rats.

Experimental variations on the rat-in-a-T-maze theme have been investigated, yielding similar results. For example, in a somewhat less controlled experimental setting, two experimenters, standing by a pond, set apart from each other at some distance, throw food to ducks at two different rates. Very quickly, the ducks are able to calculate the distinct rates of feeding, and match their time near each experimenter accordingly, spending more time at the location of greater payoff, and switching to the location of lesser payoff for a percentage of time that matches the lower yield. I should point out that these ducks' behaviour is not merely a conditioned response to a reward schedule, since they do not necessarily receive any food before matching their behaviour to the probability of payoff. Rather – and remarkably – they are able to predict the payoff before it is received! So it's clear that animals are sensitive to the *potential* probability of reward and quickly match their behaviour accordingly.

It turns out that similar statistical calculations underlie aspects of human linguistic behaviour, in that the nature and extent of variation in speech is indeed largely matched as listeners become speakers. So let's consider an example of probability matching in phonology, that is, how variation in others' productions come to be matched in one's own productions during the course of language learning. Our focus is on the word-initial stops in English that we write 'b, d, g'. All along, I have been transcribing these b̥ d̥ g̊, the hollowed circles indicating that the stop closure is mostly voiceless, with vocal fold vibration beginning just around the point when the closure is released into the next vowel. But now it won't surprise you to learn that these transcription conventions fail to capture the actual variation in these sounds' production. We find token-to-token variation, and also variation depending on the location of the stop closure itself. Typically, the farther forward in the mouth the stop closure is, the more often the closures are genuinely voiced; the farther back in the mouth the closure, the less often the closures are voiced. (We'll go into the phonetic motivation for this variation in Chapter 6.) Research on young English-learning children shows that they initially produce all their word-initial stops – whether orthographic 'p t k' (the so-called fortis category) or orthographic 'b d g' (the so-called lenis category) – something like p t k, with neither the aspiration ($\mathrm{p^h}$ $\mathrm{t^h}$ $\mathrm{k^h}$) nor minimal voicing (b̥ d̥ g̊) that is characteristic of the adult fortis and lenis categories, respectively; such young children may still lack the articulatory prowess to match the patterns they hear. Through three years of age, the two stop categories for English begin to take shape, in that some word-initial fortis stops are aspirated, but still, voicing during the stop closure is extremely infrequent in the lenis series, though less so for stops made at the lips. Even up to six years of age, children's lenis category involves fewer voiced tokens than adults'. Finally, only after six years of age do learners come to largely match the nuanced variability of their elders.

As with probability matching in lower animals, such behaviour betrays an extremely sophisticated statistical analytic ability on the part of language learners. Moreover, children's eventual productions betray evidence that they are able to implement their calculated probabilities in their own speech with startling, though imperfect, accuracy. While we can never know for sure, a rather straightforward account of probability matching in speech production might consist of speakers randomly choosing one out of their pool of stored tokens each time they speak a word. So token variants that they hear often are more likely to be chosen, and token variants that they hear less often are less likely to be chosen. In this way, the overall distribution of tokens will be well matched from speaker to speaker and from generation to generation. Speaking, then, is not like playing darts at all. Even expert dart players can't hope to accurately match the variation of their opponents.

Of course, every person's linguistic experience is different from every other person's. This is even true among individuals with very similar linguistic experience such as siblings. Consequently, if variation is largely a consequence of experience, each individual's variation will be different – in some cases ever-so-slightly different – from every other person's. But within a *speech community*, such differences – by definition – are never sufficiently great to adversely affect communicative success.

'It is not a hypothesis that children do probability matching during language learning. It is simply a description of the observed facts'. So writes William Labov in his 1994 book. The *fact* that children engage in probability matching during language learning fully supports the *hypothesis* that they also engage in exemplar modelling of variation and categorization and casts strong doubt on both the relaxed constraint approach and the prototype approach. Neither of these other approaches is properly equipped to handle the fact that variation is largely matched from generation to generation. Recall: both of these approaches view variation as created anew by each speaker, unconstrained by the extent and nature of variation to which these speakers are exposed. They, therefore, incorrectly predict that speaker-to-speaker and generation-to-generation variation will not be probability-matched. Finally, I should also point out, again echoing Kruszewski, that the proposal of an exemplar and probability matching approach is consistent with the gradual nature of sound changes. Kruszewski's remarks, though not couched in the parlance of modern cognitive science, are perhaps all the more remarkable for exactly that reason and warrant quoting at length:

> In the course of time, the sounds of a language undergo changes. The spontaneous changes of a sound depend on the gradual change of its articulation. We can pronounce a sound only when our memory retains an imprint of its articulation for us. If all our articulations of a given sound were reflected in this imprint in equal measure, and if the imprint represented an average of all these articulations, we, with this guidance, would always perform the articulation in question approximately the same way. But the most recent (in time) articulations, together with their fortuitous deviations, are retained by the memory far more forcefully that the earlier ones. Thus, negligible deviations acquire the capacity to grow progressively greater …

In other words, a prototype model does not readily allow for the possibility of gradual sound changes, since prototypes are presumably fixed. But allowing for both variation and probability matching, and differential sensitivity to recent versus remote tokens, the gradual nature of sound change may be accounted for quite straightforwardly.

But before wholeheartedly embracing the exemplar and probability matching approach, I'd like to address a plausible objection to its account of variation. Isn't it possible that the cross-generation stability of variation is not rooted in the nature of categorization but is instead purely physiological in origin? That is, since we all have comparable speech apparatus, mightn't the similar distribution of variants across the generations simply follow as a natural physical consequence? Well, yes, this is certainly a possibility, but there are a few good reasons why we should be sceptical of this explanation.

First, if variation in speech were solely a consequence of physiological pressures, then we might expect the nature and extent of variation to be nearly identical across languages with similar sound systems. For example, given two languages with similar vowel inventories, we might expect that the phonetic variation found in these languages' vowels should be extremely similar. But in fact, this doesn't seem to be the case. Both the extent and the nature of variation are different from language to language, even among those whose sound inventories are otherwise quite comparable. A similar result emerges when we investigate nasalization on vowels when followed by a nasal consonant. Every language investigated has a certain amount of variable vocalic nasalization in this context, but different languages vary in different ways. The nasalization will always be there but to different extents in different languages. So variation itself seems to be *conventionalized* on a language-specific basis. This sort of language-specific conventionalization is readily understandable under the exemplar and probability matching approach but is difficult to reconcile with a purely physiological account of speech variation.

Second, probability matching in language is found in domains that are surely not explicable in physiological terms. Some studies have shown that the optional use of certain morphemes – for example, agreement markers in certain grammatical constructions in Caribbean Spanish – is probability-matched across speakers: the rate of these morphemes' presence versus absence is *conventionalized*. For example, the Spanish plural marker is variably present on both nouns and adjectives in these dialects. We may imagine the plural marker being used 95 per cent of the time in the context where a plural meaning is intended, and so is not used 5 per cent of the time. It turns out that this usage pattern won't significantly vary from speaker to speaker but instead will be conventionalized throughout the speech community.

Third, these sorts of results have also been reproduced in the speech laboratory. In one such study, subjects were taught a contrived mini-language in which nouns were optionally marked with a made-up definite article. Subjects were divided into groups which differed in the extent to which the nouns they heard possessed this marker: one group was exposed to nouns, 75 per cent of which had the marker, and another group was exposed to nouns, 25 per cent of which had the marker. After sufficient exposure to the mini-language, subjects were asked to produce sentences in the taught language. Remarkably, subjects matched their usage to their exposure. That is, subjects in the 75 per cent group produced about 75 per cent of their nouns with the marker, and subjects in the 25 per cent group produced about 25 per cent of their nouns with the marker.

Since the exemplar and probability matching approach offers a clear and satisfying account of conventionalized morphological variation – variation which cannot possibly be attributed to physiology – there would seem good motivation to propose a similar account of conventionalized phonetic variation as well. And after all, as William Labov writes in 1994, in a discussion of probability matching in language learning,

> We should not be embarrassed if we find that systematic readjustments in … language are governed by the same cognitive faculty that governs the social behaviour of mallard ducks … We are products of evolving history, not only our own but that of the animal kingdom as a whole, and our efforts to understand language will be informed by an understanding of this continuity with other populations of socially oriented animals.

Probability matching promotes category separation and phonetic stability

Given the evidence for probability matching in language learning, it becomes quite understandable how phonological systems remain quite stable from generation to generation. Actually, I now want to argue that, in a seemingly paradoxical fashion, the excellent-though-imperfect matching of speech variation actually serves to *curtail* the very variation that is being matched! Let's consider how this can be so.

In Chapter 2, I mentioned that front vowels are usually unrounded, like i and e, and back vowels are usually rounded, like u and o. I wrote that this is probably because keeping the lips unrounded and keeping the tongue in a front position combine to create a very short oral cavity, while rounding the lips and backing the tongue combine to create a longer oral cavity. The difference in the lengths of the oral cavities corresponds to a difference in the acoustic qualities of front and back vowels. The second formant is significantly higher for front unrounded vowels and is significantly lower for back rounded vowels. These differences, I suggested, are good from a functional standpoint, because they render the different vowel qualities less confusable with one another. Of course, there are languages that *do* have front rounded vowels, as our discussions of Finnish and Hungarian vowel harmony have shown us, for example. But the overwhelming tendency is that if a language has front rounded vowels, then the language has front *un*rounded vowels as well. So Finnish and Hungarian have y, but they also have i; y's acoustic properties are somewhat intermediate between i and u, since it involves lip rounding (serving to lower F2) and tongue fronting (serving to raise F2). The idea I'm getting at here is that the vowel qualities in any given language tend to be *dispersed* in terms of their acoustic qualities. The fewer the vowels, the more distinct from each other they tend to be. Consequently, as the vowel system gets more crowded, the acoustic distinctions among the vowel qualities necessarily decrease. Compare, for example, the Spanish vowel system – quite a common one in that it contains only five members – with that of many American English dialects. These are shown in Figure 5.5.

In a five vowel system like Spanish, the vowels are symmetrically dispersed quite widely in terms of their acoustic qualities, which for our purposes include the first two formants (though there are many other phonetic

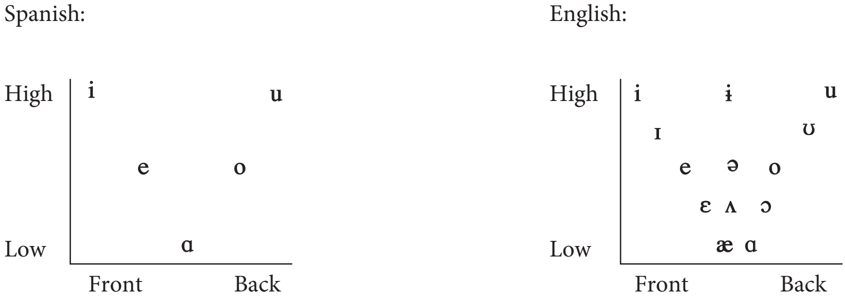

Figure 5.5 The dispersed symmetry of the Spanish and English vowel systems.

differences as well). Since English has so many more vowel qualities than Spanish does, the vowel space is more tightly packed, but still, the vowels are symmetrically dispersed, and avail themselves of a comparable overall acoustic space. In fact, we can see that the Spanish system is basically a subset of the English system in that all the Spanish vowels have acoustically similar correlates in the English system: i e ɑ o u are present in both languages.

The particular subset relation isn't the only conceivable one, however. We could well imagine that Spanish or another language might have one of the following subsets of the English system shown in Figure 5.6.

In fact, no language has anything remotely approximating these lopsided distributions. In all likelihood, it is neither by design, by intention, nor by chance, that vowel systems take the dispersed forms that they do. Rather, it is most likely due to a form of *evolution*, specifically, of *natural selection*. Vowel systems take the forms they do exactly because there are *selectional pressures* to keep vowel categories dispersed, so that words are rendered distinct from one another.

Before going even one step further, there are important aspects of the arguments I will be developing that require clarification. Specifically, when

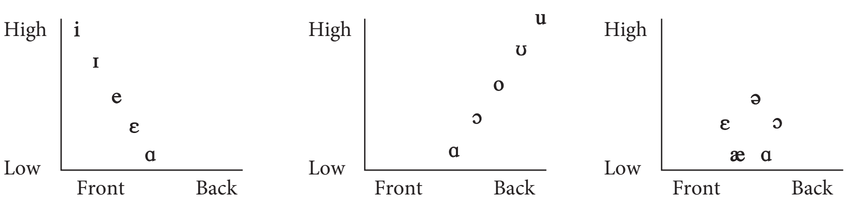

Figure 5.6 Non-existent asymmetries in vowel systems.

I say 'natural selection', what exactly do I mean, or, more to the point, what exactly do I *not* mean? First, I expressly do not mean that vowel systems, or any other structural properties of language, are genetically transmitted from generation to generation, and as such, are subject to the genuinely evolutionary pressures that genetic mutations allow. I don't mean this *at all*. Second, I am *not* proposing that the dispersion we observe in vowel systems today historically derives from vowel systems that did *not* have this quality of dispersion. There is no reason to believe that the general characteristics of vowel systems have ever been significantly different from what they are today. Third, I am not proposing a theory of evolution that allows for goal-directed behaviour on the part of the individual speaker. Indeed, Darwin's theory of natural selection does not admit this possibility. Probability matching suggests that speakers are primarily engaged with *copying* the speech patterns that they hear around them and not with *actively modifying* their speech patterns so that one sound is rendered more distinct from another.

So what *do* I mean then? If vowel systems are not genetically endowed, and if they are not a consequence of design, intention, or chance, then what is the nature of this 'natural selection' that I propose is influencing their symmetrical shape? Well, we now know that there is inherent variation in speech, such that no token is ever identical to any other token. Inevitably, tokens deviate from each other in terms of their articulatory, acoustic, and auditory properties. Nonetheless, tokens of particular vowels do cluster together – and *away* from other vowels – such that the speech signal is transmitted quite effectively to listeners, and so variation in production is well matched from generation to generation. I said 'well matched', but not 'perfectly matched'; *any* system of reproduction – genetic or otherwise – is subject to imperfect copy.

What is the locus of this imperfection? In fact, both language *perception* and language *production* are demonstrably imperfect. Most tokens, of course, are perceived accurately, in the sense that the meaning intended by speakers is recovered by listeners, because the tokens sound remarkably similar to previous tokens of the same word. These correctly perceived tokens are also usually produced as accurate copies. However, once in a while, the production of one vowel might stray a little too close to the phonetic quality of some other vowel. For example, every once in a while, a Spanish word which usually has e might be made with a somewhat higher tongue position, and end up sounding like i. Such stray tokens are inevitable; systems of reproduction are *never* perfect.

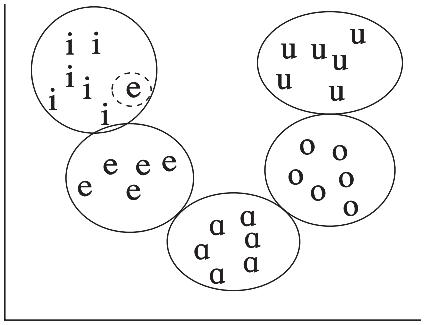

Figure 5.7 Vowel production.

With this in mind, let's reconsider the vowel inventory of Spanish, this time employing the 'cloud of tokens' notation I introduced at the beginning of this chapter, but allowing for the presence of these stray tokens, as in Figure 5.7.

Tokens situated well within a given cloud keep a safe distance from all the other vowel qualities, and thus are sufficiently distinct from vowels in the other clouds so that misinterpretation is not a problem. In all likelihood, these will be unambiguously communicated to listeners and quite accurately reproduced. The pooling together of these variable tokens into a single category is indicated, as always, by the circles. (Of course, learners do not come pre-equipped with these categories, these circles. Rather, they emerge from experience with pairing sound and meaning.) However, once in a while there will be tokens that should be grouped with one vowel quality but stray into the region of another vowel quality. Look at the stray token of e in the dashed circle. Although the word with which this token is associated almost always has a mid-front vowel, this particular token was made with a slightly higher tongue position, so that it is largely indistinguishable from words that usually have i.

As we know, listeners are usually able to overcome any ambiguities in the speech signal, because the context – real-world or grammatical – will serve to clarify meaning. So, if listeners encounter such a stray token, chances are fairly good that they nonetheless supply the word with the meaning intended by the speaker; we'll consider the consequences of these correctly interpreted strays in the next section. But learners, who are still getting the hang of pairing sound with meaning, are still developing their knowledge of the real world and their knowledge of grammar. Consequently, they are, on rare occasion, less able to recover the intended meaning of these stray tokens. Actually, it's been shown that adults too may misinterpret these stray tokens, more often than you might imagine, in fact. By my reckoning, there are at least three different ways that listeners might misinterpret this confusing Spanish token: (1) if the stray vowel quality results in another word of the language (e.g., *mella* meja 'notch' is produced as *milla* mija 'mile'), they could conceivably pair the token with the wrong meaning; (2) they might assign the token to more than one meaning; and (3) if the stray token results in a meaningless word, the token might remain uninterpreted. Each of these sorts of misinterpretation has the potential to induce confusion on the part of the listener, since the meaning intended by the speaker is not recovered by the listener. So, almost all tokens will be unambiguous, but *some* tokens will be confusing to listeners, and will remain uninterpreted or assigned to the wrong meaning.

How do stray tokens affect the probabilities that learners come to match in their own speech? Consider the pool over which learners determine the phonetic distribution of tokens. Within-pool variants are clustered together, but stray tokens – those that fall within the phonetic space of some other value, and also, ambiguous tokens that are at the outer reaches of the cluster – might be ignored, since they may not have been categorized properly. Consequently, these confusing tokens will not be pooled with the vowel quality that is normally employed for that particular word. Since these will be not be pooled with other tokens of these vowels, this results in categories consisting of distinct pools of tokens with fairly sizeable phonetic buffer regions separating them. And since listeners can only match probabilities to their *perceptions* of speakers' productions, and not to speakers' productions directly, they might conclude that the variation in the speech signal is *not as extensive* as it actually is. That is, they *overestimate* the percentage of speakers' non-stray tokens and match this estimate in their own speech.

So now let's consider how a learner might perceive the array of tokens that were produced by our Spanish speaker. Look at Figure 5.8.

The token of the e word that had strayed into i's territory has not been perceived as such by the listener. So, as these listeners become speakers, their productions – which largely match the distribution of variation that they perceive – also consist of pools of tokens with fairly sizeable buffer regions separating them (with, of course, new strays now and again). The uninterpretability of stray tokens thus actually serves to reinforce the distinctions among the categories themselves, driving one category farther away from others, and thus rendering the linguistic system more effective

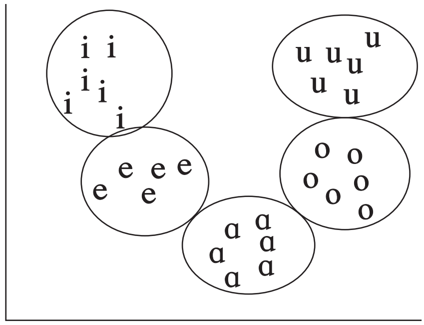

Figure 5.8 Vowel perception.

in fulfilling its communicative function. These uninterpreted tokens also serve to promote the long-term stability of the phonetic qualities of the vowel system, since usually, only tokens that are similar to the norm will be accurately perceived, and, in turn, accurately produced. So, under the exemplar and probability matching account, production errors directly *create* variation in speech but consequent perception errors indirectly *curtail* variation in speech.

Now, before going on, it's important to keep something in mind. The confusions induced by stray tokens are *not* comparable to the ambiguities of standard meaning-merging sound substitutions. Meaning-merged forms are part of the regular, fully patterned linguistic system, and so listeners encounter them all the time, and rather easily come to master their distribution. By contrast, stray tokens are not regular or patterned in their occurrence; they are genuinely aberrant, and so listeners are not equipped to deal with them in a comparable way.

Let's go back now and revisit Figure 5.1 in the beginning of this chapter. We can see now that the array of tokens in that display was not a realistic one, in that the tokens were dispersed very uniformly across the acoustic space. But we now know that, due to the uninterpretability of strays, tokens in the border regions may very well be eliminated, and probability matching will maintain the separation of categories. The revised display in Figure 5.9 reflects this more realistic distribution of tokens.

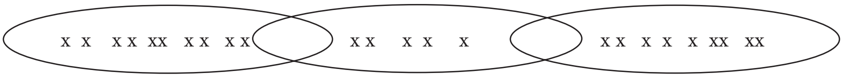

Figure 5.9 Separation of categories along a single phonetic dimension.

Of course, we will inevitably find a few strays located in the border regions, but still, the distribution of tokens across the acoustic space is probably far less regular than our first figure indicated, with most tokens falling into well-separated pools, the passive product of language *use*: the gathering of tokens towards the centre of their distributional pools is an automatic and passive consequence of the system *in use*.

To summarize, in general, speakers do a remarkably good job of matching the variability that is present in the speech signal, and listeners do a remarkably good job of perceiving this variability. However, the system isn't perfect: there are both stray tokens and consequent perception errors that influence the categorization procedure. The passive filtering-out of these strays increases the phonetic distinctions among tokens belonging to different semantic categories. The result is that vowel systems tend to avail themselves quite well of the phonetic space, dispersing their members into well-defined, well-separated regions. So the dispersion of vowel qualities in the phonetic space, and the buffer regions among them, may be seen as the natural, passive consequence of the miscommunication of stray tokens. The idea then, is that our excellent-though-imperfect ability to engage in probability matching both *promotes* and *inhibits* variation in speech. And the phonetic separation and stabilization of categories is as much a *consequence* of effective communication as it is a *cause* of effective communication.

We can now see how imperfect copy may lead to the symmetrical distributions that we observe time and time again in vowel systems. Surely, this derived symmetry should not be viewed as some sort of cognitive pressure in the minds of individual language users that favours the symmetrical distribution of elements. I don't think the symmetry of the system is relevant at any psychological level by language users; it's only appreciated by linguists. Indeed, these processes are extremely slow-acting and so cannot be attributed to individual speakers; speakers are excellent in mimicking what they hear, and so changes are very gradual. Rather, the symmetry evolves passively, as a function of language use within a community of speakers.

This sort of system may be viewed as both self-organizing, and self-sustaining. It is self-organizing because its structural properties are a consequence of its use, requiring no outside monitor, guide, or pressure, to affect its overall structure. It is self-sustaining because, by its very use, it repairs and maintains itself. So once again, language *structure* is inseparably intertwined with language *use* and language *function*.

Probability matching promotes category separation and phonetic *change*

We've just seen how imperfect copy might contribute to the phonetic separation and phonetic stability of sound categories, hence enhancing the phonetic distinctness among elements of meaning as well. However, sounds *do* change, and these changes are embodied in the slightly different distribution of tokens that are observable as the generations proceed. We'll now consider a rather different effect: imperfect copy might lead not to stabilization but to an *increased* separation of sound categories. This mechanism, in fact, is already built into the system as we have characterized it. Since tokens of one category that are more distinct from tokens of other categories are more likely to be perceived correctly, then categories may drift farther apart over the generations, but only provided that this drift does not come to encroach on the phonetic character of yet *another* category.

Given the enormously complex interaction of disparate pressures in phonological systems, asymmetric subsystems are bound to develop, at least temporarily. So let's imagine a situation in which contrastive categories, for one reason or another, are not fully dispersed in the perceptual space. Under these circumstances, one category may increase its phonetic distance from another, and no third category is present to provide a limiting counterforce. For example, we might imagine a hypothetical language like Spanish, except that it lacks a high front vowel (an admittedly unlikely system, but a fine way to illustrate the effect).

A wildly stray i-like token of a word that usually possesses e may well induce confusion on the part of listeners, since it is so different from the vowel qualities that they are used to (Figure 5.10).

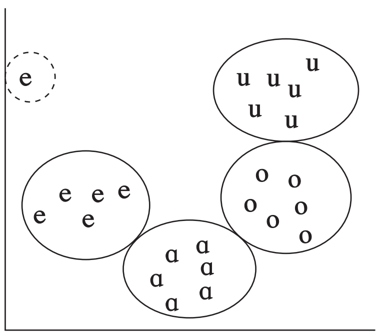

Figure 5.10 Vowel production with a 'wild stray'.

Such tokens will probably be thrown away – filtered out – regarded as mere speech errors. If noticed, they might be laughed at by both speaker and listener (Figure 5.11).

However, a token of an e-word that is only marginally i-like – not a *wild* stray, but rather, a *mild* stray – might not induce confusion at all, but on the contrary, might be better at communicating the intended message to listeners, since this token is actually further dispersed from the other vowels of the system (ɑ o u), though not outlandishly distinct from other es (Figure 5.12).

In this language, since such tokens marginally drift farther and farther away from other categories are *not* encroaching on a third category, then it's *these* tokens that are most effective in conveying the meaning intended by speakers, as displayed in Figure 5.13.

Over time, the whole pool of tokens may gradually drift farther and farther away from the other categories, and a more symmetrical four-vowel system might emerge; other values may now spread out to exploit the entirety of the available acoustic space. The result is that the system will evolve towards a symmetrical dispersion of its categories (Figure 5.14).

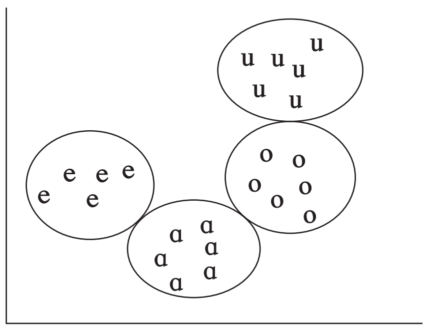

Figure 5.11 Vowel perception with the 'wild stray' filtered out.

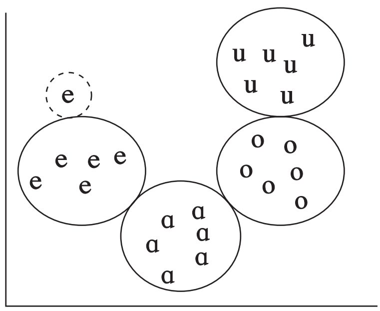

Figure 5.12 Vowel production with a 'mild stray'.

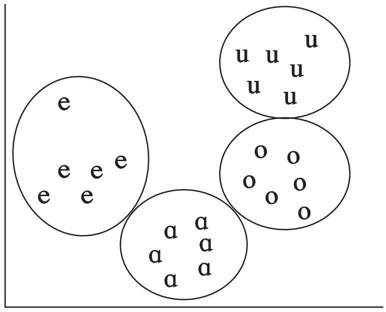

Figure 5.13 Vowel perception with the 'mild stray' pooled.

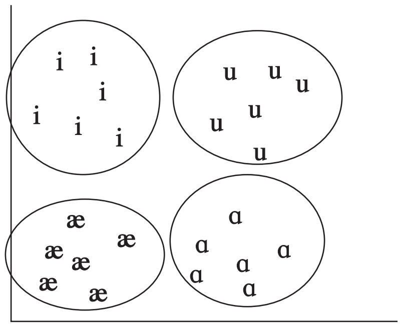

Figure 5.14 Newly evolved system.

So let's move away from the hypothetical realm, and consider a real-world example of this sort of dispersion effect.

Trique trans-velar labial harmony

Trique (pronounced 'tɹ̥ikeɹ̯, though the Spanish spelling Triqui is often seen, pronounced 'tɾiki) is a language spoken in the southern regions of Mexico. It is a member of the Otomanguean language family. In Trique, whenever a round vowel precedes a velar consonant (k g), it is immediately followed by w. So look at the examples in Table 5.1; I've underlined the relevant spans.

You'll notice that when we have u followed by either k or g, there is a w immediately following: the language does not have sequences such as ukɑ or ugɑ. You can readily see how the w is merely a continuation of the u,

Table 5.1 Round vowels and velars in Trique

nukwah	Strong	dukwa	possessed house
dugwah	to twist	zugwi	(name)
ʒugwa	to be twisted	dugwe	to weep
dugwane	to bathe (someone)	rugwi	peach
rugwah	hearth stones	dugwi	together with, companion

such that the lip-rounding gesture is realized both immediately before and immediately after the consonant. The pattern, then, may be conceived of as a minor form of vowel harmony.

It wasn't always this way, however. Employing the comparative method, when we investigate other languages that are closely related to Trique, we do *not* observe the 'unhinging' of round vowels. Related languages usually have uka and uga in words for which Trique has ukwa and ugwa. And since less common patterns among closely related languages often reflect more recent changes, Trique was probably innovative in this sense. Evidence from the internal reconstruction method supports the Trique pattern as an innovative one, as these ws are completely predictable in their presence such that we can 'undo' their present distribution and recover an earlier stage of the sound pattern. Consequently, both the comparative and the internal reconstruction methods converge on the same conclusion about the history of Trique: at an earlier stage of the language, the round vowel was *not* realized on both sides of the velar consonants. Instead, there was *uka and *uga, patterns that are completely absent today. (Recall that reconstructed forms are traditionally indicated with an asterisk preceding them.) But at some point in the language's history, round vowels began to 'unhinge' from their position, and continue across velar consonants – eventually becoming ukwa and ugwa, respectively.

But now look at the set of words in Table 5.2. Here – when the consonant that follows the round vowel is alveolar – the w is *not* present. In fact, it's *never* present here. So, we never find utwa or udwa, for example.

The question that a phonologist must now ask is, why did the Trique pattern arise? Why did u harmonize across velars k g but not across alveolars t d? The answer I'd like to pursue is that this minor form of vowel harmony *enhances* the acoustic distinction between the velar and alveolar consonants. Recall that alveolar consonants like d bring F2 towards about 1800 Hz as the closure is being released. By contrast, when releasing a velar consonant like g into a vowel, F2 begins at about 1600 Hz. (We'll just be discussing d and g

Table 5.2 Round vowels and alveolars in Trique

rune	large black beans	utɑh	to anoint
utʃe	to get wet	utʃi	to nurse
utɑ	to gather	dunɑ	to leave something
gunɑh	to run	rudɑʔɑ	stone rolling pin
ʒutʃe	hens, domestic fowl	gunɨ	to hear

from here on out, but all arguments apply to t and k as well.) This means that the difference in F2 between, say, dɑ and gɑ is about 200 Hz at consonantal release. (There are several other acoustic differences between these two sounds, and so they are not terribly likely to be confused with one another.) Now, if the velar consonant is altered such that a w is superimposed onto its release, the oral cavity becomes longer, and so F2 lowers. In fact, F2 lowers rather significantly, to about 900 Hz. This means that the difference in F2 between dɑ and gwɑ is about 900 Hz (1800 Hz minus 900Hz). Clearly, the superimposition of the w increases the acoustic distance between the release quality of alveolar and velar consonants. Importantly, since gw and kw sequences were elsewhere absent in the earlier stage of the language, harmonizing lip rounding across g increased the acoustic distinction between these consonants and the alveolar consonants, *without* encroaching on another sound category. So ugɑ could become ugwɑ, and since there were no other words in the language like ugwɑ, there was no passive semantic counter-pressure acting to inhibit the sound change.

Harmonizing across the alveolar consonants, by contrast, would serve to *diminish* the velar – alveolar acoustic distinction. Why is this so? Superimposing a w onto the release of an alveolar consonant would change the F2 onset from about 1800 Hz to about 1500 Hz, decreasing the difference in F2s to a mere 100 Hz (1600 Hz minus 1500 Hz). So, an accompanying change from udɑ-to-udwɑ would have undone the functional benefits of the ugɑ-to-ugwɑ change, as portrayed in Figure 5.15.

Of course, this w didn't just pop out of the ether in order to help increase the acoustic distinction between udɑ and ugɑ, as our transcriptional system might suggest. So, for example, we wouldn't expect an s or an m to arise in order to enhance the perceptual distinctness in the ugɑ context. Instead, these sorts of changes exploit the sounds that are already 'loitering in the neighbourhood', so to speak, their properties harnessed, co-opted, or, in the parlance of Stephen Jay Gould and modern evolutionary biologists, *exapted* to fulfil new functional roles: gɑ and dɑ are not especially

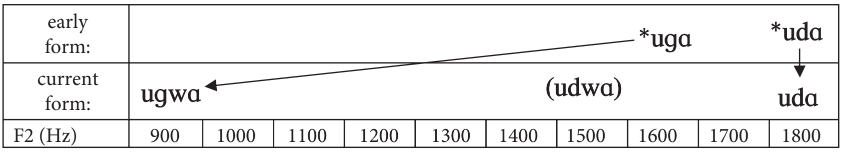

Figure 5.15 In Trique, the diachronic harmonizing of labiality onto the release of the velar consonants (ugwa) increased their F2 distinctions with the alveolar consonants (uda) and didn't encroach on the perceptual space of another category. Harmonizing labiality onto the release of the alveolar consonants (udwa) would have had counter-functional consequences.

confusable with each other, but since u was right next door, and since its harmonizing across g served only to increase the acoustic separation of the elements without jeopardizing semantic clarity, there was nothing to inhibit the beneficial change. So u was passively recruited to enhance the acoustic distinctness among otherwise less distinct forms. Over time, the number of ugwa variants was likely to increase, since these forms increased the acoustic distance from uda, and so were more likely communicated correctly to listeners. Meanwhile, uda remained largely stable over time: udwa-like variants were confusable with both uga and the increasing number of ugwa tokens, and so were not likely to take hold. The proposal, then, is that due to the acoustic and consequent functional advantages of harmonizing labiality across the velar consonants, and the disadvantages of harmonizing across alveolar consonants, the Trique system evolved towards its present state.

Now, let's consider how the exemplar and probability matching approach may account for sound changes like this. First, of course, there is inherent gradience and variation in speech production, and so uga ~ ug̜a ~ ugwa and uda ~ ud̜a ~ udwa are among the possible variants that any speaker might produce. (The subscripted hook indicates partial rounding; variation, unlike alternation, is indicated with the wavy dash.) In the earlier stages of the language, productions leaned heavily towards uga and uda, just as they still do in the languages related to Trique. However, stray ugwa-like variants rendered these words *more* distinct from their uda counterparts. This is especially true since words with ugwa were *not* previously present in the language. Consequently, there was no counter-pressure inhibiting a change towards ugwa. Therefore, those variants with w were more likely communicated unambiguously to listeners. Meanwhile, ambiguous tokens were sometimes confusing to listeners. Specifically, udwa-like variants of words that usually had uda may be confused with ugwa, and so weren't added to the pool of tokens over which probabilities were calculated; they were passively ‘repelled’ due to the presence of ugwa forms. Consequently, as the generations proceeded, listeners were more likely to perceive ugwa and uda as unambiguously phonetically – hence semantically – distinct, and so they were more likely to produce ugwa and uda in their own speech, as a passive consequence of probability matching.

So, the variation engaged in by elders was largely matched by learners, but nonetheless, due to the greater likelihood of unambiguous perception of certain variants over others – ugwa over uga; uda over udwa – learners’ calculated probabilities may have differed slightly from their elders’, in that the variants that were more dispersed from the opposing value were more often perceived correctly, and so, in turn, more often produced. In essence, the presence of ambiguous tokens may result in listeners *overestimating* the prevalence of more distinct tokens. This overestimation, in turn, may passively enhance phonetic distinctness among forms, hence also enhancing the semantic clarity of the speech signal.

These proposals are summarized in the Figure 5.16, which demonstrates how very minor phonetic tendencies, coupled with the confusion they might induce or eschew, may eventually have far-reaching consequences for the sound system.

Let’s consider this in a bit more detail. Look at Figure 5.17. Entering the sound change midstream, we might take a 1000 token sample from one generation of speakers. Let’s call them Generation W. Of these tokens, 750 are uga, while 250 are ugwa. Most of these tokens are produced as a consequence of learners’ matching their probability of occurrence to the productions of Generation V. In turn, Generation X perceives *all* ugwa tokens unambiguously. Among uga tokens however, let’s suppose that a

*uga~uģa ~ugwa		*uda~uḑa~udwa	
↙	↘	↙	↘
less distinct from uda	more distinct from uda	more distinct from uga	less distinct from uga
↓	↓	↓	↓
less likely perceived unambiguously	more likely perceived unambiguously	more likely perceived unambiguously	less likely perceived unambiguously
↓	↓	↓	↓
less likely produced	more likely produced	more likely produced	less likely produced
∴ gradual move towards ugwa		∴ stability of uda	

Figure 5.16 The fates of *uga and *uda.

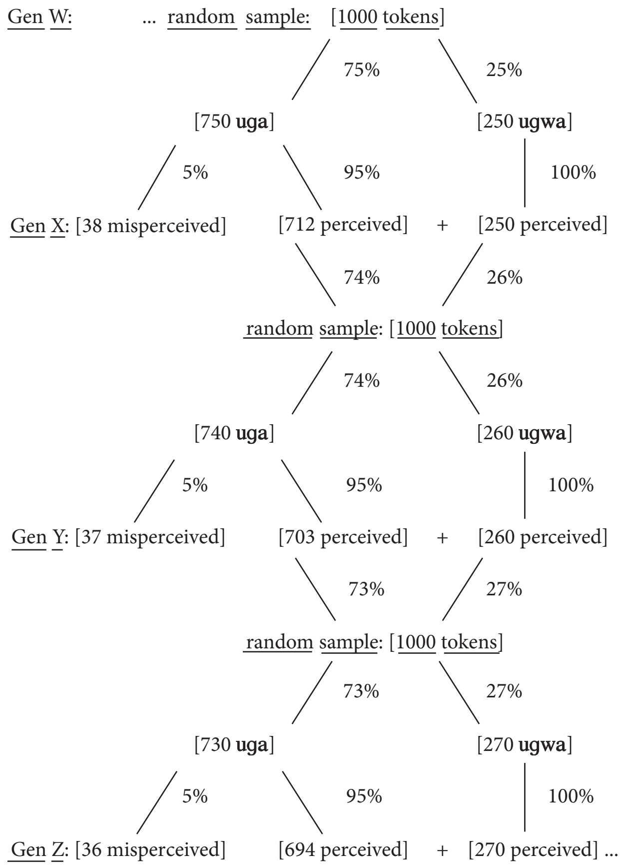

Figure 5.17 Schematic diachrony of uga-to-ugwa.

full 5 per cent of these 750 tokens (38 in all) are confusing to listeners, since their acoustic separation from uda is not as sharp. These 38 misperceived tokens will not be pooled with those over which Gen X-ers calculate their probabilities. Now we iterate the process: if we take a random sample of 1000 of Generation X's *productions*, we should observe that they largely match the probabilities that they *perceive* their elders to have produced. Generation X perceived 712 out of 962 tokens as uga (38 tokens were misperceived); this constitutes a rate of 74 per cent. So, out of 1000 tokens produced by Generation X, 740 will be uga, and 260 will be ugwa. And again assuming that 5 per cent of the uga tokens will be misperceived by Generation *Y*, *these* children will perceive only 73 per cent tokens as uga (703 of 963 tokens), and so on down the generations. We may now see, given the small tendency to better perceive ugwa tokens, how, over the course of time, the conventions of the language may change: language, though in a constant state of flux, inevitably maintains semantic clarity.

A model like this does not perfectly or exhaustively predict specific language patterns: *we can no better predict the future direction of a sound pattern than we can the future direction of a species*. Indeed, one of the best advantages of this account is that it effectively captures the *probabilistic* nature of sound change. Trique's relations did not undergo the sound change that Trique did. There simply exists some level of likelihood that any given sound change will take hold in any given language. Probabilities may be affected by, among many other factors, the pre-existing language-specific sound pattern: in Trique, the introduction of labialized velars was contrast-enhancing, since spreading could not induce homophony; a meaning-maintaining sound change. In some other language, a prevalence of contrastive labialized velars might very well passively induce the curtailment of such a sound change.

These sorts of proposals for the origin and evolution of sound changes may actually be reproduced in a laboratory setting. A laboratory condition may serve to recapitulate elements of the hypothesized historical scenario in 'speeded-up' form if we find a way of inducing a high rate of perception errors on the part of listeners. How might we do this? I had subjects listen to uda, udwa uga ugwa in various levels of 'white noise' (computer-generated noise across a broad frequency range which decreases the signal-to-noise ratio, thus making the signal harder to decipher). Noise introduced into the speech signal might induce a 'speeded-up' rate of misperception in certain contexts, and thus reflect one origin of real-world sound change. I found, indeed, that listeners were far more likely to hear uda as uga than they were uda as ugwa. Among the four forms presented, these latter two forms (uda and ugwa) were the least often confused with one another; udwa and uga were the most often confused with one another.

Now, this sort of result doesn't immediately translate into a real-world context that unfolds over generations of speakers, but nonetheless, it is noteworthy – and consistent with my proposed explanation – that in my experiment, the least confusable forms (uda and ugwa) are exactly those which actually seem to have evolved in Trique from more confusable forms (uda and uga). So, given that language learners largely (though imperfectly) match the variation they perceive, the sorts of perceptual errors induced in my experiment might only reflect the culmination of a slow, generation-to-generation accretion of such errors, rather than offering any major insights into the online processing of natural speech. Nonetheless, the results are consistent with the hypothesis that the gradience and variation inherent in speech production may be the fodder for these sorts of sound changes: the more distinct the variant from an acoustically similar word, the more likely that it will be interpreted correctly, and so the more likely the system will wend towards this value.

The past is not only reflected in extant phonological alternations; the same sorts of pressures that gave rise to the present state of the system may, at least in theory, be brought to the fore under the proper laboratory conditions, not necessarily by modifying the natural speech *signal* (recall, I have expressed strong reservations about putting much stock in listeners' interpretation of artificially modified speech) but instead modifying the *noise* that accompanies this signal. As remarked by Baudouin de Courtenay in 1910, 'I must emphasise the importance of errors in hearing (*lapsus auris*) when one word is mistaken for another, as a factor of change at any given moment of linguistic intercourse and in the history of language as a social phenomenon. Experimental methods can help to define the types and directions of these errors …'.

Before concluding our discussion of Trique, there is an important point to consider. In both my discussion of the actual Trique system, and in the experiment I performed, I have been operating under the assumption that uga and uda constituted the critical distinction between the words that drove the sound change. However, it's simply not the case that a huge inventory of Trique words was originally differentiated *solely* in terms of whether they had uda or uga. Usually, words with these sequences had additional elements that rendered them distinct, such as the presence of word-initial consonants (e.g., utah 'to gather' versus nukwah 'strong', which have the voiceless stop counterparts), and/or different tones (I haven't been indicating tones, but Trique is a tonal language). Moreover, if some words were indeed solely differentiated by uda versus uga, couldn't Trique have evolved the ugwa pattern only in those specific cases in which homophony might otherwise be the result, while individual words *not* at such a risk remained unchanged?

Indeed, it may be that ugwa in Trique first arose in those very uga words that were minimally distinct from uda words, that is, in those words that were identical except for their g versus d. But these few pioneering uga words that evolved towards ugwa may have opened the floodgates of change: as *some* words were now implemented with ugwa, more and more words may have quickly fallen in line with the emerging pattern. Why might this have happened? Due to the pioneering ugwa words, the language now possessed three relevant patterns: uda, uga, and ugwa. Of these three patterns, uda and uga are phonetically much more similar to each other than either is to ugwa. At this point, when an uga word was now heard, it was more likely to be confused with uda than with the newly developed ugwa words; since ugwa has now entered the language, new stray ugwa variants were more likely to be recognized, and so were communicated more effectively to listeners. We might even say that the new presence of new ugwa words *attracted* uga words towards them, though I want to emphasize again that any such attraction was not due to the intentions – conscious or otherwise – on the part of speakers. As always, *language change is a passive consequence of language use.*

In sum, semantic pressures may have triggered the change to ugwa in some words. But now that ugwa was present in the system, it was far more likely that additional uga words would fall in line with the new pattern, since uga is far more confusable with uda than it is with ugwa. Figure 5.18 portrays the proposal in schematic form.

We actually observe this sort of scenario time and time again in phonology. The linguist Joan Bybee, for example, has demonstrated that sound changes often begin in a word here, a word there, but eventually come to permeate the language. Mirroring the proposals of nineteenth-century scholar Hugo Schuchardt, Bybee observes that it is the most frequently employed words that might change first (more on this in Chapter 7), but what I'm suggesting here is that *passive pressures towards homophone avoidance may also trigger individual words to undergo pioneering changes.*

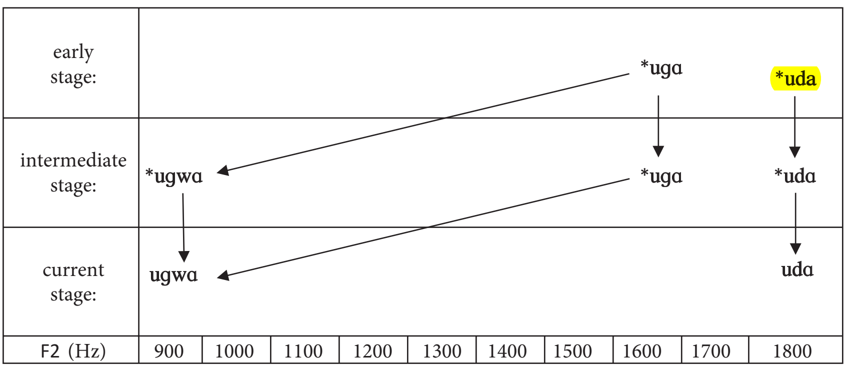

Figure 5.18 The proposed Trique change: the first words to become *ugwa may have been *uga words that minimally contrasted with *uda words. With these *ugwa forms now in place, the pattern was more likely to generalize, changing all *uga words to ugwa. Meanwhile, *uda remained stable.

To summarize, the conventions established by speech communities betray a nuanced mastery of the phonetic variation internalized by individual speakers that is demonstrably a part of these speakers' linguistic knowledge. The exquisite articulatory control that speakers display in their productions is best evidenced by the fact that they are able to largely match in their own speech the variation present in the ambient pattern. On this view, learners' articulatory talents may be harnessed largely in service to *copying* or *imitating*-not *modifying* (improving upon or otherwise)-the ambient speech pattern. But still, speakers' mimetic talents are not perfect. Stray tokens are inevitable, and it is the functional benefit of certain of these strays that might ultimately take hold and come to permeate the system.

Comaltepec Chinantec tone alternation

Like a classical Darwinian approach to the evolution of species, I've just suggested that the origin of labial harmony in Trique is rooted in two related phenomena. First, random, minor inexactitudes of speech production slowly amass over generations of speakers, such that one generation's inexactitudes serve as the next generation's template for copy. The result is that variation is largely – though imperfectly – matched over generations of speakers. Second, beneficial variants are more likely to be perceived correctly by listeners, and so it's *these* variants that are more likely to survive and propagate as listeners become speakers. These beneficial phonetic variants may come to be generalized throughout the language.

In Trique, change was initiated by purely random, directionless, phonetically *isotropic* chance, since variation potentially proceeds in a radially symmetrical fashion. Variation may have proceeded in any direction – maybe a little *less* rounding, maybe a little *more* rounding, maybe a little bit of something else – but s*ome* tokens just happen to have better functional success over others, and so the sound change moved in that direction. Maintaining rounded lips through a velar consonant is no more phonetically natural than *not* maintaining this lip rounding, but it is due to the functional advantages of labial harmony that the Trique sound system began its new trajectory.

In this section, I'd like to consider a slight variation on the theme exemplified by Trique. Some sound changes, although also (and inevitably) subject to the sorts of semantic pressures discussed for the Trique pattern,

9781474238885_txt_rev.indd 180 27/03/17 10:17 PM

are actually 'helped along' by certain natural phonetic tendencies. What I mean is, certain variants may be more likely than others due to purely phonetic pressures. And if these variants are *semantically* beneficial as well, then a sound change is more likely to be channelled in that direction. The variation which leads to sound change in this scenario is not phonetically *isotropic* but is instead phonetically *anisotropic*.

Chinantec, like Trique, is a member of the Otomanguean language group. The Chinantec dialect we are interested in is spoken in the beautiful mountainside village of Santiago Comaltepec (koˌmɑlteˈpɛk, a four-hour bus ride north of the city of Oaxaca, Mexico. The Comaltepec dialect of Chinantec, like all Otomanguean languages, is tonal. Comaltepec Chinantec words may have a low tone (˩), mid tone (˧), high tone (˥), low-to-mid tone (˩˧), or low-to-high tone (˩˥), along with a few alternants that we'll discuss momentarily.

In Comaltepec Chinantec, there is a rather complicated tone substitution pattern, aspects of which we'll be considering now. First, when a low-high (˩˥) tone comes to precede a form that otherwise has a low tone, then high-low (˥˩) is found instead of low (˩). As always, elements in alternation are underlined in Table 5.3.

Second, when a low-high (˩˥) tone precedes a form that otherwise has mid (˧), then high-mid (˥˧) is found instead of mid, as in Table 5.4.

Finally, when a low-high (˩˥) tone follows a form that otherwise also has low-high (˩˥) tone, this latter tone is substituted with mid-high (˧˥), as in Table 5.5.

Table 5.3 Low (˩) changes to high-low (˥˩) after high (˥)

to:˩	banana	kwɑ˩˥ to:˥˩	give a banana
ŋɨh˩	chayote	kwɑ˩˥ ŋɨh˥˩	give a chayote

Table 5.4 Mid (˧) changes to high-mid (˥˧) after low-high (˩˥)

ku:˧	money	kwɑ˩˥ ku:˥˧	give money
ndʒœ:˧	jug	kwɑ˩˥ ndʒœ:˥˧	give a jug

Table 5.5 Low-high (˩˥) changes to mid-high (˧˥) after low-high (˩˥)

ʔŋɑ˩˥	forest	he:h˩˥ ʔŋɑ˧˥	in the forest
mbʌʔ˩˥	ball	kuɑ˩˥ mbʌʔ˧˥	give the ball!

Now, there are two interesting generalizations we can make about these sound substitutions, one generalization about their phonetic character, and one about their semantic character. Phonetically, we can characterize the sound substitution in much the same way we did the lip-rounding harmony process of Trique. Specifically, the high component of the high-low (˥˩) and high-mid (˥˧) contour tones may be viewed as being a mere extension of the preceding high tone (˥), moving across the intervening consonant and continuing into the first part of the next vowel. So the substitution of high-low (˥˩) for low (˩), and high-mid (˥˧) for mid (˧) is a consequence of the preceding high tone (˥) being implemented both before and after the intervening consonant. The mid-high tone (˧˥) may be viewed in similar terms, the preceding high tone (˥) serving to at least partially raise the first portion of the following low-high tone (˩˥).

The second interesting generalization is a semantic one. Recall that I've listed five tone values for Comaltepec Chinantec: ˩ ˧ ˥ ˩˧ ˩˥. Any of these five tones may occur on forms that do *not* immediately follow words with low-high (˩˥) tones. However, we've just discussed three more tones that may *only* occur when following low-high tones: high-low (˥˩), high-mid (˥˧), and mid-high (˧˥) (high and low-mid may occur here as well, but low, mid, and low-high do not). In other words, (1) low alternates with high-low, (2) mid alternates with high-mid, and (3) low-high alternates with mid-high, *all in a meaning-maintaining fashion*. Set notation is provided in Figure 5.19.

Given these two generalizations, we're now in a position to understand the origins – the *explanation* – for this aspect of the Comaltepec Chinantec sound system. The first point to consider is a phonetic one. It's been shown experimentally that pitch rises take longer to implement than do pitch falls. This is schematically portrayed in Figure 5.20.

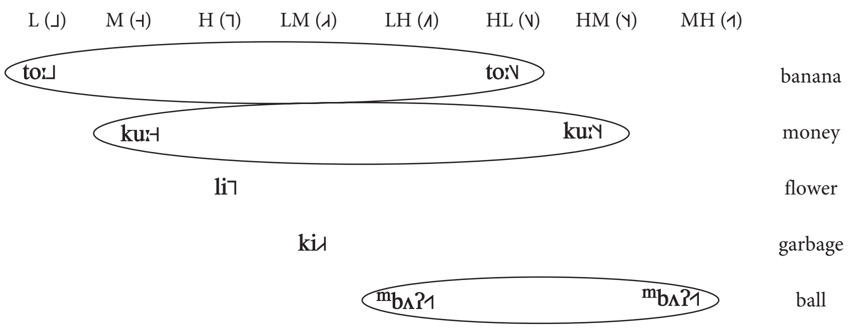

Figure 5.19 The meaning-maintaining nature of Comaltepec Chinantec tone substitution.

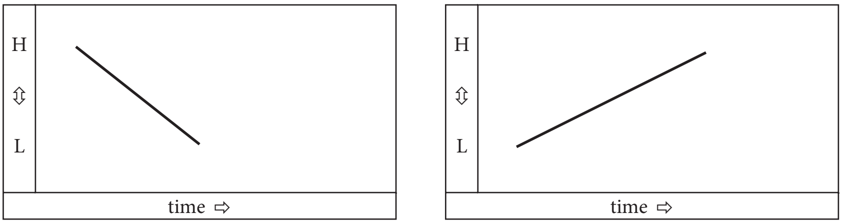

Figure 5.20 Pitch rises take longer to implement than do pitch falls.

Given the sluggishness of pitch rises in comparison to pitch falls, a consonant may already be made *before* a pitch rise is fully achieved: upon the release of this subsequent consonant, finally, maximum pitch height is achieved on the next vowel. The idea then is that, rising tones are more likely than falling tones to 'spill over' their high component on to a following vowel. Since falling tones can be produced faster than rising tones, they might be less likely to spill over onto the next vowel. In Figure 5.21, I've superimposed consonants ('C') and vowels ('V') on the pitch patterns. The potential for high 'spillover' should be clear.

Comaltepec Chinantec has conventionalized this phonetic tendency in a way that fairly hugs the physical limitations of the speech apparatus. The high component of low-high contour tones is implemented both at the end of the first vowel and into the beginning of the second vowel.

But still, just because speakers' physiological limits might be encountered in an experimental context doesn't mean that these limits will play a role in natural linguistic contexts. Indeed, only if it can be shown that speech patterns *exactly match* experimentally determined physical limitations can we establish a direct link between phonetic limitations and phonological patterning. In fact, as far as I know, an *exact* match between physiological constraints and linguistic conventions has *never* been established in linguistic research. For example, it's been found that women can raise their pitch more quickly than can men, but no language is sensitive to such sex-based

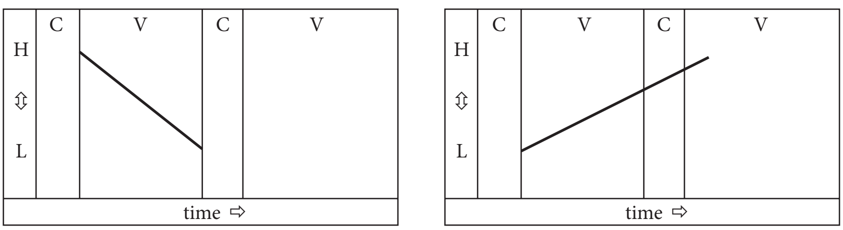

Figure 5.21 *High*-tone 'spillover' from rising tones.

differences. (There are, in fact, some languages for which men and women have slightly different sound inventories, but these differences are arbitrary, in the sense that they do not exploit inherent physiological differences between men and women.) Nonetheless, physiological constraints might come to constrain phonological patterning at a historic distance. That is, the conventions of sound systems might not push the absolute limits of physiology, but might nonetheless come to be diachronically shaped by them.

This is where semantic pressures on the system become relevant, which may, over generations of speakers, crucially interact with phonetic pressures. If the high end of the low-high contour did not spill over in Comaltepec Chinantec, then the resultant pitch pattern might be misperceived by the listener as belonging to the low-mid tone category, due to the only limited temporal domain in which the pitch rise is implemented. The pitch rise may be cut off as the second consonant is beginning and so does not achieve nearly as high a value. That is, with its lower ending point, the tone might be confusable with low-mid tones, at times thus culminating in a meaning-merging sound pattern. This scenario is schematized in Figure 5.22.

As greater pitch increases can be effected *after* the following consonant, and since there is a natural tendency for pitch rises to 'spill over' anyway, the low-high tone may be better cued when it spills over. Also, since high-low, high-mid, and mid-high are elsewhere absent in the language, high-tone spillover better conveys the high-tone value without the possibility of meaning-merger. So, variant forms in which the high component of low-high forms spills over on to a following vowel are semantically advantageous in two ways: (1) The low-high tone is now less likely to be confusable with low-mid tones, and (2) the spilled-over component can never merge meanings, since high-low, high-mid, and mid-high are purely meaning-maintaining sound substitutions in Comaltepec.

What I'm suggesting is that high-tone spillover has its origins in phonetically anisotropic variation: there is an intrinsic phonetic pressure

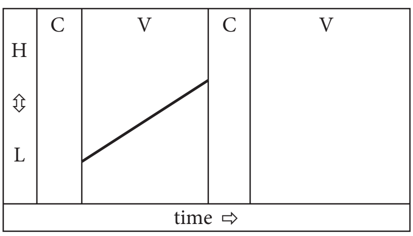

Figure 5.22 Early cut-off of a *low-high* tone may be confusable with a *low-mid* tone.

towards on high-tone spillover in Comaltepec Chinantec. And exactly because the high-tone spillover has *semantic* value – in that distinct words are now more readily successfully conveyed to listeners – this tone value has been conventionalized in its present form.

There are complications, however: mid tones on syllables that lack post-vocalic **h** or **ʔ** induce the same tone substitutions on following syllables as do low-high tones, as exemplified in Table 5.6

Also, these meaning-maintaining mid-high and high-mid tones may themselves induce tone changes on following syllables! Since they induce the tone change when they are low-high and mid, they also do so even when they have been substituted by mid-high and high-mid, respectively. That is, the tone substitution pattern may iterate itself across a span. In theory then, a series of forms may possess nice smooth mid tones in most contexts, but suddenly encounter a very bumpy road if they end up next to each other: ˧, ˧, ˧, ˧, but ˧ ˥˧ ˥˧ ˥˧!

This unusual pattern calls out for an explanation. Why should a high tone suddenly pop up after a mid tone? Clearly, the proposed explanation cannot offer an immediate account of this pattern. Indeed, I'm really at a loss to offer any sort of proximate, phonetically rooted account here. But of course, proximate phonetic pressures are only of limited help when trying to fully understand synchronic phonological patterns. Instead, the answer may be recoverable from an investigation of language history.

As in Trique, both the comparative and internal reconstruction methods converge to offer a compelling account of the present-day pattern. Forgoing the details, the linguist Calvin Rensch has reconstructed an earlier stage of Chinantec, and suggests that the mid tones which induce a high tone on following syllables are historically derived from high tones, as presented in Table 5.7.

Table 5.6 Low (˩) changes to high-low (˥˩), mid (˧) changes to high-mid (˥˧), and low-high (˩˥) changes to mid-high (˧˥) after mid that lacks a following **h** or **ʔ**

hi˩	book	mɨː˧ hi˥˩	I ask for a book
mohʔ˩	squash	mɨː˧ mohʔ˥˩	I ask for squash
kuː˧	money	mɨː˧ kuː˥˧	I ask for money
ʔoː˧	papaya	mɨː˧ ʔoː˥˧	I ask for papaya
ŋi˩˥	salt	mɨː˧ ŋi˧˥	I ask for salt
loh˩˥	cactus	mɨː˧ loh˧˥	I ask for a cactus

Table 5.7 Reconstructed high tones in Chinantec, and their modern incarnation as mid tones

Historic	Present-day	
Chinantec	Comaltepec Chinantec	
*kuː˥	kuː˧	money
*dʒuː˥	ndʒœː˧	earthen jar/jug
*ʔwɨː˥	ʔwɨːŋ˧	Ojitlán (a large village)

Table 5.8 Reconstructed high tones in Chinantec, and their modern incarnation, unchanged

Historic	Present-day	
Chinantec	Comaltepec Chinantec	
*lih˥	lih˥	flower
*huːh˥	huːh˥	word
*huːʔ˥	huːhʔ˥	pineapple

Historic high tones that were immediately followed by h or ʔ have remained high, according to Rensch, as in Table 5.8.

If this is so, the origins of this superficially strange pattern can be suggested. We have established that the high component of low-high tones naturally spills onto a following vowel. Over time, this pattern may have *generalized* to include other tones that ended on a high pitch. In particular, high tones on vowels that were not immediately followed by a laryngeal sound (h or ʔ) were the most likely tones to be recruited into the pattern.

Why should this be so? As we've already discussed, glottal opening (h) and glottal closing (ʔ) typically make demands on the vocal folds that are in conflict with pitch production. If a laryngeal sound immediately follows the high tone, the vocal folds are not likely to maintain the posture necessary for the production of this tone. *Without* a following laryngeal sound however, a pitch may be prolonged into the following vowel without interference. So, level high tones also came to be associated with the appearance of a high tone on the first portion of a following vowel, provided no h or ʔ immediately followed. But then, these high tones lowered to mid, and so all these historic high-tone forms disappeared from the language. There are reasonable aerodynamic reasons for this sort of pitch differential, since both shutting down and opening up the glottis may both – for rather different reasons – be accompanied by pitch rising: **h** involves increased airflow, which might raise pitch (recall our discussion of Seoul Korean in Chapter 4), and **ʔ** involves tensed vocal folds, which also might raise pitch. In Comaltepec Chinantec, these historic high tones are lower today, but the following high tone remains, as a relic, or vestige, of the past pattern. So whenever one of these former high tones (which had now become mid) came up against a following low, mid, or low-high tone, these following tones were still substituted with high-low, high-mid, and mid-high, respectively: the preceding high tone has lowered to mid, but the substitution pattern on the following vowel remains solidly in place (somewhat comparable to Hungarian's pattern of historically back-unround roots retaining their back suffixes even after having fronted).

As a result of all this, former high-tone words that *lacked* a post-vocalic laryngeal are now mid-tone words, and they induce the presence of a high tone on a following vowel. But the high tones that *possessed* a post-vocalic laryngeal remained high, yet these do *not* spill their high on to the following vowel! The timeline in Figure 5.23 summarizes the proposed sequence of changes.

Time 1	Time 2	Time 3	
What: H spreads rightward from LH syllables.	What: H spreads rightward from H-final vowels which lack postvocalic **h** or **ʔ**.	What: Level H *without* postvocalic **h** or **ʔ** lowers to M; the H tone on the following vowel remains.	What: Level H *with* postvocalic **h** or **ʔ** remain H; there is no H-tone spread
Why: Functionally beneficial anisotropic variation leads to the conventionalization of H-tone spill over in this context.	Why: The pattern is generalized to include those H-final vowels most susceptible to spill over: those lacking **h** or **ʔ**.	Why: Lack of post-vocalic laryngeals lead to a phonetically natural pitch split, while the allophonic substitution remains unchanged.	Why: Presence of post-vocalic laryngeals (**h** and **ʔ**) make demands on the vocal folds that may be in conflict with tone production.
Example: **toː˩** 'banana' **kwɑ˩˥ toː˥˩** 'Give a banana'	Example: **hi˩** 'book' **mɨː˧ hi˥˩** 'I ask for a book'	Example: **kuː˧** 'money' **mɨː˧ kuː˥˩** 'I ask for money'	Example: **ʔneh˥** 'need' **ʔneh˥ ni˩ kih˧** 'We need to pay'

Figure 5.23 Proposed timeline of Comaltepec Chinantec tonal alternations.

A final complication involves low-mid tones, which are never substituted with other tones when low-high or mid immediately precedes. Why not? Well, if, in the history of Comaltepec Chinantec, high tones were to slowly 'spill over' on to low-mid tones, then low-mid might be substituted with high-mid. The problem is that high-mid tones are already found in this context, as alternants of mid tones. In other words, if low-mid were to alternate with high-mid here, it would run the risk of ending up identical with mid tones that alternate with high-mid here. Although we really don't know the historical sequence of events that led to this fortuitous 'blocking' of a potentially meaning-merging alternation, suffice to say for now, the present-day pattern may be remarkably well-motivated when the considering semantic pressures that result in communicative success.

Some readers may feel a sense of dissatisfaction at my liberal use of these sorts of conjectures, and indeed, it would be very nice if linguists could state with certainty the evolutionary origins of present-day phonological patterns. But of course, we don't have phonetic and semantic data across centuries of language use. Nonetheless, I do hope to have demonstrated that linguists, even with their limited data and limited tools, might arrive at plausible – if not necessarily confirmable – conclusions about a language's previous state.

To summarize, the present-day tone patterning in Comaltepec Chinantec, as superficially strange as it is, can be understood as the culmination of a series of small, local, and emphatically *natural* incremental changes. First, phonetically anisotropic variation of low-high-tone 'spillover' may have become conventionalized, since those low-high tones which, quite naturally, spilled over onto a following vowel, were better at keeping words phonetically distinct from each other that differed in meaning. These variants were thus naturally selected and eventually conventionalized. The pattern seems to have generalized to include level high tones as well, but only those that lacked post-vocalic laryngeals. Without conflicting demands placed on the vocal folds (due to the absence of a following h or ʔ), it was *these* tones that were most naturally incorporated into the pattern. And although these tones subsequently lowered to mid for rather well-understood aerodynamic reasons, the tone-change process had been fully conventionalized by this time, and so we still observe – up to today – a high tone on the first portion of the following vowel. The result is that this aspect of the pattern cannot be synchronically motivated, but so what!

Phonological patterns evolve in the absence of the users who ultimately come to master them.

The appearance of high tones following mid tones in Comaltepec Chinantec exemplifies something quite remarkable about the nature of sound substitution. Even when a phonological pattern seems to be downright bizarre, lacking any reasonable phonetic or semantic motivation at all, there will be perfectly natural, incremental processes that have unfolded over time that may account for the pattern. A series of small, local, interactions of phonetic and semantic pressures may, over generations of speakers, render alternants quite distinct from each other: *alternations in the present – even when phonetically unnatural and superficially counter-functional – are the long-term product of small, local, and perfectly natural phonetic and semantic pressures that play themselves out over generations of speakers.*

Still, phonological systems tend to remain remarkably natural and phonetically plausible, even though the ravages of time logically allow for bizarre patterns to slowly emerge. This disparity between reality and logical possibility can be reconciled quite intuitively, however: with every utterance by every speaker in every language, phonetic reality exerts pressure on the system. As unnatural patterns slowly emerge (always, of course, as a consequence of slow, natural, and local steps), phonetic pressures will always be exerting themselves, due to the simple fact that each speech utterance is an actual physical event that unfolds in real time and so is subject to genuine physical pressures. In time, irregular, unusual patterns may once again be slowly shaped by raw physiology. Consequently, phonetically implausible patterns are constantly under pressure to fall back in line, and so those that do survive are not only phonetic oddities, but are statistical oddities as well.

In the case of Comaltepec Chinantec at least, we may have successfully uncovered some of the major pressures on the system that have led to its present state. But even in those cases when the present state of the system is hopelessly obscured by long-forgotten, undocumented historical changes (Taiwanese tonal behaviour immediately comes to mind, for example), we should not just throw up our hands and give up the notion that *all* phonological patterns are explainable by real-world pressures that are known to act locally in any number of natural circumstances. The optimistic nature of scientific pursuit demands us to operate under the assumption that broadly applicable, locally active principles go far in explaining the complex world around us, in phonology, and elsewhere.

High-tone behaviour in Zulu and elsewhere

As mentioned in the previous section, high-tone 'spillover' is a rather common pattern cross-linguistically. Indeed, it's well established that high tones are more phonologically 'active' than non-high tones, in terms of their 'shiftability', among other things. This is an obvious consequence of the phonetic properties that set them apart from other tone values, in particular, the greater amount of time it takes to reach a higher pitch, as opposed to the time it takes to reach a lower one. Zulu, for example, displays a high-tone shift in the context of an immediately preceding 'depressor' consonant, a class of consonants that is characterized as phonetically and/or historically *breathy-voiced*, that is, involving both vocal fold vibration (at one end of the vocal folds) and vocal fold spreading (at their other end). Depressors acquired their evocative name from the pitch lowering that accompanies them upon consonantal release. We've already seen in our discussion of Korean stop-tone alternation how vocal fold vibration during a stop closure serves to lower pitch upon the stop's release into a vowel. Now, in the case of a breathy-voiced release – which, recall, is further accompanied by a significant lowering of the entire larynx – the rate of vocal fold vibration is lowered even more, culminating in an especially lower pitch upon its release. (These pitch perturbations are very brief, and we rarely have conscious awareness of them. At any rate, we'll motivate the co-occurrence of larynx lowering in this context and other, similar ones in Chapter 7.) In Zulu, a depressor that is immediately followed by a short high-toned vowel realizes the pitch peak on the *following* vowel (schematically, **CV˥CV˩**, but **DV˩CV˥˩**; where **D** = depressor, **C** = non-depressor). However, displacement is blocked if *another* depressor immediately follows (**DV˩DV˩**;). Also, it does not take place from long vowels. Instead, the pitch rise is implemented on the vowel itself.

For example, both z and ŋ̤ are depressors, as they possess both voicing and vocal fold spreading (the subscript x̤ indicates breathiness). Consequently, when the following vowel is short and high toned, these consonants are accompanied by pitch lowering on their immediately following vowels, and the presence of a high tone on the following vowel: i˩si˥ɬaː˩lo˩ 'chair' – i˩zi˩ɬaː˥˩lo˩ 'chairs'; i˥ɲoː˩nɨ˩ 'bird' – ŋ̤e˩ɲoː˥˩nɨ˩ 'with a bird'. Meanwhile, there is no high-tone displacement if the vowel is long: ziː˩˥˩kʰoː˩na˥ 'they being present' – zi˩kʰoː˥˩na˥ 'they are present'.

It's easy to see the parallels between the phonetic situation in Zulu and that in Comaltepec. In both cases, the pattern involves a low pitch that is immediately followed by a high pitch on the same vowel. Again, since pitch rises take longer to implement than do pitch falls, then in the context of a Zulu depressor consonant, a following high tone may be achieved only *after* the following consonant has been completed, culminating in an apparent displacement of the high tone. Moreover, in Zulu, just as in Comaltepec Chinantec, conflicting laryngeal gestures may block the propagation of the process; indeed, it should not be surprising that an immediately following depressor blocks tone displacement, since the vocal folds return to their breathy-voiced – hence pitch-lowering – state. In this context, it might not surprise you to learn that the 'trapped' higher pitch is actually accompanied by a slightly longer vowel, so that it can 'do its best' to be realized saliently, such that all distinguishing phonetic information is encoded in the speech signal: **D<u>V˩</u>CV˥˩**, but **D<u>Vˑ˩˥</u>DV**˩ (*not* **D<u>V˩</u>DV**˩, as previously schematized; moderate lengthening is notated with ˑX). My jocular anthropomorphism here is not really a laughing matter though, and may actually be a bit depressing: as we'll explore in Chapter 7, there is a school of phonological thought that firmly believes in *goal-directed behaviour* on the part of speakers, a linguistic 'itch' that is, by this school's assertions, buried deep in the human genome; not funny at all!

Instead, of course, the longer vowel may have *passively evolved*, simply because *slight* vowel elongation here did a better job of cueing semantic distinctions to listeners, and thus took hold; *slight*, because *significant* lengthening would create a new set of problems for poor Zulu: since the language possesses a vowel length contrast, if short vowels after a depressor were fully long here (say, **DVː˩˥DV**˩ rather than **DVˑ˩˥DV**˩), they would clearly run the risk of inducing homophony with minimally distinct forms that possess a long first vowel, whether a depressor follows or not.

Given both the phonetic and semantic pressures that interact in both Comaltepec and Zulu, it should not be surprising that quite a number of other languages display similar patterns of high-pitch spread/displacement. In Beijing Mandarin, rising tones typically peak only after the following consonant has been implemented; tones with low offsets show a significantly lesser spillover effect in these same contexts. In Zagreb Croatian, pitch contrasts involve a rising pitch contour on one vowel, with the pitch peak being realized on the next vowel. Both Peninsular and Mexican Spanish have very comparable patterns, in that stressed syllables typically possess rising pitch, again, the pitch peak being realized after any following consonant is released in the next vowel. Especially because Spanish and Croatian lack a rich and crowded system of tonal phonology (i.e., they are otherwise non-tonal), their high tones are rather free to spill over without affecting the auditory perspicuity of any other tones, and so are rather free to evolve in a phonetically natural way, that is, to spill over on to the following vowel.

It is exactly due to such complex diachronic interactions between phonetic and semantic pressures on the communicative system that, it turns out, high tones are typically more 'phonologically active' than low tones, in that high tones tend to move around, whereas low tones tend to be more stable, hugging more dearly to their lexically affiliated morphemes.

Summary and conclusion

In this chapter, we've explored in some detail how variation in speech can sometimes lead to confusion for listeners, and how this confusion may ultimately lead to an increased phonetic distance among semantically distinct forms. We've seen how, under some circumstances, variation may induce the phonetic stability of categories, but under other circumstances, variation may induce sound change. Under both sets of circumstances, however, I attributed the variation inherent in speech production to the accumulation of minor, chance errors over generations of speakers. Again, *sounds in alternation in the present, which undergo quantum leaps of change in phonetic quality as they shift from context to context, have passively evolved in the absence of the users who come to master them*.

Alternations may now be viewed as the culmination of a series of small, natural changes to the system that take place over generations of speakers. Indeed, even when a pattern does not lend itself to a compelling explanation in the present, we should not abandon the idea that interacting phonetic and semantic pressures are ultimately responsible for its linguistic behaviour. *Present-day alternations have no present-day causes; they only have present-day effects*.

The regularities we observe in sound change and in alternations are merely the most easily observed consequences of the genuine laws and principles that underlie linguistic sound patterns, laws and principles that cannot be readily observed, since they do not have overt, unique analogues in linguistic patterns. As Baudouin de Courtenay wrote in 1910,

> Many scholars, who are either undemanding or incapable of critical thinking, confuse law, that is, functional interdependence, with statistical statements of fact or with plain coincidence. Others posit logical, methodological, and epistemological axioms, set up conditions sine qua non for each scientific proposition, and formulate subjective laws for any theoretical ideas in place of objective laws that account for the relationships of observable facts.

Baudouin de Courtenay wrote that the *genuine* law-governed primitives that operate on linguistic patterns derive from four main sources: (1) 'the external, physical world' ('sound'), (2) 'the psychological world of the individual' ('mind'), (3) 'the biological and physiological world of a given organism,' ('body'), and of course, (4) 'the social world (the transmission of linguistically expressed ideas from one individual to another'. Operating under the incorrect assumption that the emergent outputs of these interacting systems are directly law-governed elements themselves, Baudouin de Courtenay wrote,

> can be compared to such 'laws' as meteorological generalizations or to various kinds of statistical generalizations; in fact, they are only statements of what occurs on the surface of phenomena. Genuine 'laws', the laws of causality, are hidden in the depth, the intricate combination of the most diverse elements. 'Laws' do exist, but not where they are being sought… Between the starting and ending point of historical change… there is no relationship that could be interpreted as a law of evolution.

Doing phonology: Chiquihuitlan Mazatec

Like Chinantec, Mazatec is an Otomanguean language spoken in the southern reaches of Mexico. It should thus not be surprising if we find that the language possesses certain phonological similarities to Chinantec. Specifically, we'll consider some data from the Chiquihuitlan (ˌtʃikiˈwitlan) dialect of Mazatec that has a tonal pattern somewhat akin to Comaltepec Chinantec's: rising and higher tones induce a pitch increase on the following vowel. Indeed, as pointed out in this chapter, comparable patterns are found not only in the Otomanguean group but also in a diversity of unrelated languages. This is not surprising since, recall, a phonetically natural pattern such as this – exactly because of its naturalness – tends to be cross-linguistically prevalent. So consider the data in (5a), in which alternants (along with their associated vowels, which phonetically manifest the relevant tones) are underlined.

Unconditioned alternant:		Conditioned alternant:	
rki^{3}	medicine	sua^{1}rki^{1}	I-give-medicine
ŋka^{3}	again	kih^{31}ŋka^{1}mu^{1}su^{14}	went-again-hired-worker
mu^{3}su^{14}	hired worker	”	
hbæ^{3}	it finishes	ŋku^{2}ʔɲu^{2}hbæ^{2}	rapidly-it-finishes
nta^{3}ʔŋka^{34}	corncrib	ŋku^{2}nta^{2}ʔŋka^{24}	one-corncrib
nta^{3}ʔŋka^{34}	corncrib	ŋku^{2}nta^{2}ʔŋka^{24}	one-corncrib
kw̥a^{4}	word	ho^{1}kw̥a^{14}	two-word
ne^{4}	‘uh’	nu^{2}ne^{24}	‘uh’-year
me^{4}	they	kwiʔ^{42}me^{24}	will-drink-they
“		koh^{3}me^{34}	with-they

Figure (5a)

Meanwhile, none of the forms in (5b) is in an alternating context. That is, none of these forms is tonally conditioned.

ŋki^{3}hɲa^{2}	field
ka^{2}be^{14}tsẽʔ^{3}	began-I
ʔa^{4}kwa^{4}t^{h}ĩ^{2}	just-like-that
ka^{3}ba^{3}k^{h}ã^{2}	broke-(3p)
ʃi^{4}ntʃa^{1}	will-be-put-in-(3p)
ku^{4}ma^{3}si^{3}ne^{1}	will-become-yellow

Figure (5b)

Here, numbers indicate pitch values, **1** being highest pitch, **4** being lowest. For example, **14** is a precipitously falling pitch; **34** involves a shallow fall in pitch. Now, inspection of the IPA chart will reveal that these number symbols are not in evidence: they are not IPA symbols. Rather, these numbers are the tone symbols traditionally employed by Americanist linguists, used before the IPA rightfully reigned supreme. I earlier emphasized the importance of adhering to IPA standards in phonology, so that readers might effortlessly decipher the phonetic correlates of the written form, but please bear with me; this lapse in the IPA employed herein is more technological in origin than it is inherent to the IPA itself. The IPA has the *potential* to represent any tonal value, but available computer fonts are typically remiss in providing *all*-attested tonal values: given the various complex tones found in Chiquihuitlan Mazatec for example, some tones are missing from the IPA as supplied by most font packages. Consequently, we depart – this time only, and with my sincerest apologies – from IPA convention.

Our first job is to isolate the tones that alternate; as stated, I have underlined these (and their affiliated vowels, which phonetically bear the pitches). In most contexts (the unconditioned alternants), tones are realized in one way, but in particular phonological circumstances (the 'conditioned alternants'), these tones are substituted with others. Abstracting away from consonant and vowel values (which, I tell you now, have nothing to do with the tone alternations in Chiquihuitlan), we might quickly observe that under rather circumscribed conditions, tones of lower pitch values are substituted with tones that bear a higher pitch at their beginning. For example, the lower tone **3** in **rki**^{3} is substituted with a higher tone 1 (rki^{1}). Comparably a tone that begins with a lower tone **3**, as in ʔŋka^{34}, is substituted with a tone value that begins with a higher tone **2** (ʔŋka^{24}).

So, while the general pattern of alternation is quite clear, as phonologists we are still obliged to effectively argue for the phonetic and semantic (i.e., diachronic) origins of the pattern, in an effort to genuinely *explain* why the pattern is the way it is. Indeed, when solving phonology problems, a superficial description of the pattern is never enough. Rather, answers must be *motivated* at least to the extent that presented data make such motivation possible. We turn to such a systematic presentation now.

Let's exhaustively extract the pitch patterns of the alternants in evidence, in order to set the relevant phonetic patterns in relief. Both unconditioned tones and their conditioned alternants are culled in (5c).

Unconditioned tone	Conditioned tone
3	**1**
3	1
3	1
3	2
3	2
34	24
4	14
4	24
4	24
4	34

Figure (5c)

In each case, a lower tone is substituted with, or in the case of the lowest tone (**4**), preceded by, a higher tone in the relevant context. Such observations help us to isolate the nature of the alternations: *lower tones are replaced with or (in the case of tone* 4*) preceded by higher tones.*

But under what circumstances? That is, what has triggered, or conditioned, the tone substitution? Given our experience with Chinantec and comparable systems, a promising strategy would be to narrow our focus to the phonetic properties of the *preceding* tone. So let's do that. Let's extract these presumed tonal triggers to see if any sort of pattern emerges. In (5d), I've extracted the preceding tone so that the pattern may be readily eyeballed.

Preceding tone (presumed triggers)
1
31
1
2
2
2
1
2
42
3

Figure (5d)

With the exception of tone **3**, all preceding tones are in the upper-pitch range when placed in the context of the entirety of the attested pitch span. The exceptional case will be considered momentarily, but for now, notice the overall pattern: conditioned alternants are higher in pitch, as are their immediately preceding tones. This becomes quite clear if we now juxtapose the preceding tones (our presumed triggers of tone substitution) with their immediately following conditioned alternants, as in (5e).

Preceding tone	Conditioned tone
1	**1**
31	**1**
1	**1**
2	**2**
2	**2**
2	**24**
1	**14**
2	**24**
42	**24**
3	**34**

Figure (5e)

As may now be seen, when we juxtapose the conditioned tones with their preceding tones, it becomes patently clear that the alternation is a consequence of extending the preceding higher tone – never a lower tone – onto the following vowel, such that it either completely overwrites this second vowel's tone, or (solely in the case of tone **4**) is added to create a falling tone.

Now, we might simply state that tones are realized both on their vowel of lexical origin as well as their immediately following vowel, as an initial cursory eyeballing of the data might reveal. However, this account makes no mention of the fact that only *higher* tones precede such they spill across (thus also accounting for the otherwise anomalous lower-toned **3**-tone spillover: this lower tone raises the pitch at the beginning of the still-lower **4** tone that immediately follows).

Now, given our discussion in this chapter, the phonetic underpinnings of Chiquihuitlan tonal behaviour is no great mystery, but what of its semantic underpinnings? Does the tone change adversely affect communicative success by creating excessive homophony, thus listener confusion? The short answer is, 'of course not'. But the longer answer must acknowledge that tone spillover may indeed culminate in some homophonic forms, forms that are distinct in prevalent contexts may be rendered homophonous in a conditioned context. Such induced homophony would simply require that two forms that are identical apart from their unconditioned tonal vowels be found in a conditioning context that renders their tones identical as well.

It should be emphasized that the Chiquihuitlan Mazatec tonal phonology is, in actuality, far more complex than is suggested by the data we have considered. But none of these complexities is sufficient to override the general tendency that higher tones extend their domain of expression to encompass not only their vowels of lexical origin, but, in running speech, to propagate on to following vowels, thus raising these vowels' pitches.

In sum, the case of Chiquihuitlan Mazatec is similar in kind – though different in detail – to patterns found in other languages. This is not surprising. The plethora of pressures acting upon linguistic systems is reflected in linguistic diversity itself. Physical and cognitive pressures acting in this contingent fashion culminate a mere 'family resemblance' among the world's phonological system.

Further reading

The Martinet quote ('Linguistic identity …'):

Martinet, André (1975 [1988]). 'The internal conditioning of phonological changes', *La Linguistique* 24.2: 17–26, 25.

The Paul quotes ('However much … movement …'; 'Variability of production …'):

Paul, Hermann (1890 [1970]). *Principles of the History of Language*. College Park: McGrath, 43, 44.

The Kruszewski quotes ('Our characteristic, unconscious memory …'; 'In the course of time …'):

Kruszewski, Mikołaj (1883 [1995]). 'An outline of linguistic science', in Konrad Koerner, editor, *Writings in General Linguistics*. Amsterdam Classics in Linguistics 11. Amsterdam: John Benjamins Publishing Company, 66, 85.

On rats and ducks:

Gallistel, Randy (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.

Harper, D.G.C. (1982). 'Comparative foraging in mallards: Ideal free ducks', *Animal Behaviour* 30: 575–584.

The Labov quote ('It is not a hypothesis …'):

Labov, William (1994). *Principles of Linguistic Change, Vol. 1: Internal Factors*. Oxford: Blackwell, 583.

Probability matching in Caribbean Spanish:

Poplack, Shana (1980a). 'Deletion and disambiguation in Puerto Rican Spanish', *Language* 56.2: 371–385.

Poplack, Shana (1980b). 'The notion of the plural in Puerto Rican Spanish: competing constraints on (s) deletion', in William Labov, editor, *Locating Language in Time and Space*. New York: Academic Press, 55–67.

Probability matching in the laboratory:

Hudson, Carla L. and Newport, Elissa L. (1999). 'Creolization: Could adults really have done it all?' in Annabel Greenhill, Heather Littlefield and Cheryl Tano, editors, *Proceedings of the Boston University Conference on Language Development, 23*. Somerville, MA: Cascadilla Press, 265–276.

On the acquisition of English stops consonants:

Preston, M.S. and G. Yeni-Komshian (1967). 'Studies on the development of stop consonants produced during the second year of life', *Annual Report, Neurocommunications Laboratory* 3: 211–222. Johns Hopkins University School of Medicine.

On speech variation; on exemplar modeling, and so on:

Gluck, Mark A. and Bower, Gordon H. (1988). 'Evaluating an adaptive network model of human learning', *Journal of Memory and Language* 27: 166–195.

Goldinger, Stephen D. (1997). 'Words and voices: Perception and production in an episodic lexicon', in Keith Johnson and John W. Mullennix, editors, *Talker Variability in Speech Processing*. New York: Academic Press, 33–56.

Goldinger, Stephen D. (1998) 'Echoes of echoes? An episodic theory of lexical access', *Psychological Review* 105.2: 251–271.

Janda, Richard D. and Joseph, Brian D. (2001). 'Reconsidering the canons of sound change: Towards a Big Bang theory', Manuscript of paper presented at the International Conference on Historical Linguistics, Melbourne.

Johnson, Keith (1997). 'Speech perception without speaker normalization: An exemplar model', In Keith Johnson and John W. Mullennix, editors, *Talker Variability in Speech Processing*. New York: Academic Press, 143–165.

Liberman, Mark (2002). 'Simple models for emergence of a shared vocabulary', Paper presented at *Laboratory Phonology VIII*. New Haven, CT.

Lotto, Andrew (2000). 'Language acquisition as complex category formation', *Phonetica* 57: 189–196.

Ohala, John J. (1981). 'The listener as a source of sound change', Papers from a Parasession on Language and Behavior, Chicago Linguistics Society, 178–203.

Ohala, John J. (1989). 'Sound change is drawn from a pool of synchronic variation', in Leiv E. Breivik and Ernst H. Jahr, editors, *Language Change: Contributions to the Study of its Causes*. Berlin: Mouton de Gruyter, 173–198.

Steels, Luc (2000). 'Language as a complex adaptive system', in Marc Schoenauer, Kalyanmoy Deb, Guenter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo, Hans-Paul Schwefel, editors, *Lecture Notes on Computer Science. Parallel Problem Solving from Nature*. PPSN-VI, 17–26.

On Trique:

Hollenbach, Barbara E. (1977). 'Phonetic versus phonemic correspondence in two Trique dialects', in William R. Merrifield, editor, *Studies in Otomanguean Phonology*. Dallas: Summer Institute of Linguistics.

Longacre, Robert E. (1957). *Proto-Mixtecan*. International Journal of American Linguistics 23.4.III. Indiana University Research Center in Anthropology, Folklore and Linguistics. Bloomington: Indiana University.

Longacre, Robert E. (1962). 'Amplification of Gudschinsky's Proto-Popolocan-Mixtecan', *International Journal of American Linguistics* 28: 227–242.

Longacre, Robert E. and René Millon (1961). 'Proto-Mixtecan and Proto-Amuzgo-Mixtecan Vocabularies: A preliminary cultural analysis', *Anthropological Linguistics* 4.3: 1–44.

Silverman, Daniel (2006). 'The diachrony of labiality in Trique, and the functional relevance of gradience and variation', in Stephen Anderson, Louis Goldstein and Catherine Best, editors, *Papers in Laboratory Phonology VIII*, Mouton de Gruyter.

The Baudouin de Courtenay quotes ('I must emphasise the importance …'; 'Many scholars …'; '… can be compared …'):

Baudouin de Courtenay, Jan (1910 [1972]). 'Phonetic laws', in Edward Stankiewicz, editor, *A Baudouin de Courtenay Reader*. Bloomington: Indiana University Press, 267, 272, 276.

On sound patterns suggesting the passive curtailment of potentially homophone-inducing sound changes:

Manuel, Sharon (1990). 'The role of contrast in limiting vowel-to-vowel coarticulation in different languages', *Journal of the Acoustical Society of America* 88: 1286–1298.
Manuel, Sharon (1999). 'Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts', in William J. Hardcastle and Nigel Hewlett, editors, *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press, 179–198.
Öhman, Sven E.G. (1966). 'Coarticulation in VCV utterances: spectrographic measurements', *Journal of the Acoustical Society of America* 39: 151–168.

On homophone avoidance:

Casenhiser, Devin (2005). 'Children's resistance to homophony: an experimental study of pseudohomonyms', *Journal of Child Language* 32: 1–25.
Charles-Luce, Jan (1993). 'The effects of semantic content on voicing neutralization', *Phonetica* 50: 28–43.
Kaplan, Abby, Andrew Wedel and Scott Jackson (2013a). Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts. *Language and Speech* 56: 395–417.
Kaplan, Abby, Andrew Wedel and Scott Jackson (2013b). High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128: 179–186.
Kiparsky, Paul (1972). 'Explanation in phonology', In S. Peters, editor, *Goals of Linguistic Theory*. Englewood Cliffs, NJ: Prentice-Hall, 189–227.

On Chinantec:

Anderson, Judi L. (1989). *Comaltepec Chinantec Syntax*. Studies in Chinantec Languages v. 3. Dallas: Summer Institute of Linguistics.
Anderson, Judi L., Martinez, I.H. and Pace, Wanda (1990). 'Comaltepec Chinantec tone', in William R. Merrifield and Calvin

R. Rensch, editors, *Syllables, Tone, and Verb Paradigms*. Studies in Chinantec languages v.4. Dallas: Summer Institute of Linguistics, 3–20.

Pace, Wanda J. (1990). 'Comaltepec Chinantec verb inflection', in William R. Merrifield and Calvin R. Rensch., editors, *Syllables, Tone, and Verb Paradigms*. Studies in Chinantec languages v.4. Dallas: Summer Institute of Linguistics, 21–62.

Rensch, Calvin R. (1968). *Proto Chinantec Phonology*. Papeles de la Chinantla VI. Seria Cientifica 10. Mexico: Museo Nacional de Antropologia.

Rensch, Calvin R. (1976). Comparative Otomanguean Phonology. Bloomington: Indiana University.

Silverman, Daniel (1997a). 'Laryngeal complexity in Otomanguean vowels', *Phonology* 14.2: 235–261.

Silverman, Daniel (1997b). 'Tone sandhi in Comaltepec Chinantec', *Language* 73.3: 473–492.

On the sluggishness of pitch rises in comparison to pitch falls:

Ohala, John J. and Ewan, William G. (1973). 'Speed of pitch change', *Journal of the Acoustical Society of America* (abstract) 53: 345.

Sundberg, Johan (1979). 'Maximum speed of pitch changes in singers and untrained subjects', *Journal of Phonetics* 7.2: 71–79.

Xu, Yi, and Sun, Xuejing (2002). 'Maximum speed of pitch change and how it may relate to speech', *Journal of the Acoustical Society of America* 111: 1399–1413.

On Chiquihuitlan Mazatec:

Jamieson, Allan R. (1977). 'Chiquihuitlan Mazatec phonology', in William R. Merrifield, editor, *Otomanguean Phonology*. Dallas, TX: Summer Institute of Linguistics, 93–105.

Jamieson, Allan R. (1977). 'Chiquihuitlan Mazatec tone', in William R. Merrifield, editor, *Otomanguean Phonology*. Dallas, TX: Summer Institute of Linguistics, 107–136.